PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N | A2

(11) International Publication Number: **WO 99/54436**

(43) International Publication Date: 28 October 1999 (28.10.99)

(21) International Application Number: PCT/US99/08522

(22) International Filing Date: 16 April 1999 (16.04.99)

(30) Priority Data:

09/062,580	17 April 1998 (17.04.98)	US
09/074,984	8 May 1998 (08.05.98)	US
09/084,303	26 May 1998 (26.05.98)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	09/084,303 (CIP)
Filed on	26 May 1998 (26.05.98)
US	09/062,580 (CIP)
Filed on	17 April 1998 (17.04.98)
US	09/074,984 (CIP)
Filed on	8 May 1998 (08.05.98)

(71) Applicant *(for all designated States except US)*: EXELIXIS PHARMACEUTICALS, INC. [US/US]; 260 Littlefield Avenue, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: HOMBURGER, Sheila, A. [US/US]; 3974 25th Street, San Francisco, CA 94114 (US). PLATT, Darren, M. [US/US]; 929 Pine Street, San Francisco, CA 94108 (US). FERGUSON, Kimberly, C. [US/US]; 1038 Chula Vista Avenue, Burlingame, CA 94010 (US). DOBERSTEIN, Stephen, K. [US/US]; 2 Townsend Street, San Francisco, CA 94107 (US). BUCHMAN, Andrew, R. [US/US]; 3119 Eton Avenue, Berkeley, CA 94705 (US). REDDY, Bindu, P. [IN/US]; 855 Bounty Drive, Foster City, CA 94404 (US).

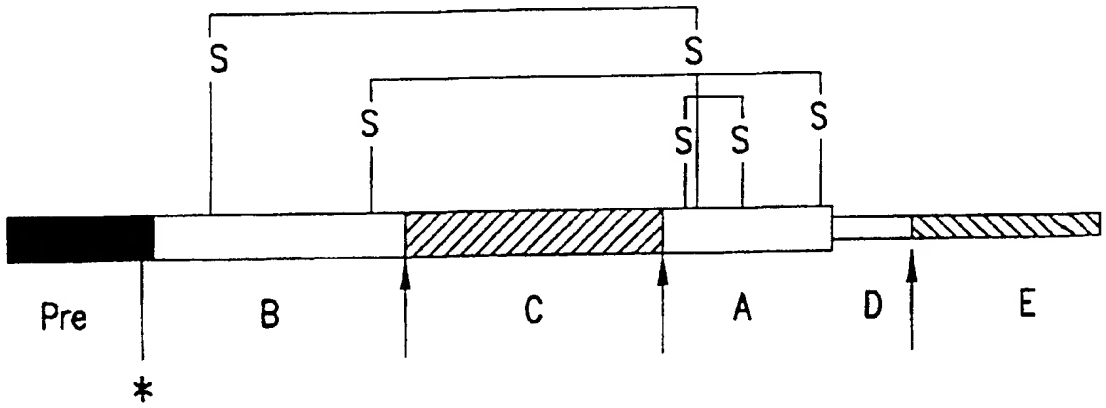
(74) Agents: ANTLER, Adriane, M. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: NUCLEIC ACIDS AND PROTEINS OF *C. ELEGANS* INSULIN-LIKE GENES AND USES THEREOF



(57) Abstract

The present invention relates to *C. elegans* insulin-like genes and methods for identifying insulin-like genes. The methods provide nucleotide sequences of *C. elegans* insuline-like genes, amino acid sequences of their encoded proteins, and derivatives (e.g., fragments) and analogs thereof. The invention further relates to fragments (and derivatives and analogs thereof) of insuline-like proteins which comprise one or more domains of an insulin-like protein. Antibodies to an insuline-like protein, and derivatives and analogs thereof, are provided. Methods of production of an insuline-like protein (e.g., by recombinant means), and derivatives and analogs thereof, are provided. Further, methods to identify the biological function of a *C. elegans* insulin-like gene are provided, including various methods for the functional modification (e.g., overexpression, underexpression, mutation, knock-out) of one or more genes silmutaneously. Still further, methods to identify a *C. elegans* gene which modifies the function of, and/or functions in a downstream pathway from, an insulin-like gene are provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NUCLEIC ACIDS AND PROTEINS OF *C. ELEGANS* INSULIN-LIKE GENES AND USES THEREOF

This application is a continuation-in-part of copending U.S. application Serial No. 09/084,303, filed May 26, 1998 which is a continuation-in-part of U.S. application Serial No. 09/074,984, filed May 8, 1998 which is a continuation-in-part of U.S. application Serial No. 09/062,580, filed April 17, 1998, each of which is incorporated by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to *C. elegans* insulin-like genes and methods for identifying insulin-like genes. The methods provide nucleotide sequences of *C. elegans* insulin-like genes, amino acid sequences of their encoded proteins, and derivatives (*e.g.*, fragments) and analogs thereof. The invention further relates to fragments (and derivatives and analogs thereof) of insulin-like proteins which comprise one or more domains of an insulin-like protein. Antibodies to an insulin-like protein, and derivatives and analogs thereof, are provided. Methods of production of an insulin-like protein (*e.g.*, by recombinant means), and derivatives and analogs thereof, are provided. Methods to identify the biological function of a *C. elegans* insulin-like gene are provided, including various methods for the functional modification (*e.g.*, overexpression, underexpression, mutation, knock-out) of one gene, or of two or more genes simultaneously. Methods to identify a *C. elegans* gene which modifies the function of, and/or functions in a downstream pathway from, an insulin-like gene are provided.

BACKGROUND OF THE INVENTION

Insulin-like proteins are a large and widely-distributed group of structurally-related peptide hormones that have pivotal roles in controlling animal growth, development, reproduction, and metabolism. At least five different subfamilies of insulin-like proteins have been identified in vertebrates, represented by insulin, insulin-like growth factor (IGF), relaxin, relaxin-like factor (RLF), and placentin (also known as early placenta insulin-like peptide, or ELIP).

Insulin superfamily members in invertebrates have been less extensively analyzed than in vertebrates, but a number of different subgroups have been defined including molluscan insulin-related peptides (MIP-I to MIP-VII) (Smit et al., 1988, Nature 331:535-538; Smit et al., 1995, Neuroscience 70:589-596), the bombyxins of lepidoptera (Kondo et

al., 1996, J. Mol. Biol. 259:926-937), and the locust insulin-related peptide (LIRP) (Lagueux et al., 1990, Eur. J. Biochem. 187:249-254). More recently, putative orthologs of both vertebrate insulin and IGF have been identified in a tunicate (McRory and Sherwood, 1997, DNA and Cell Biology 116:939-949). This is of significance since tunicates are thought to be the closest living invertebrate relative to the progenitor from which vertebrates evolved.

Apparent homologs of the insulin receptor have been identified in both the fruit fly and the nematode (Petruzzelli et al., 1986, Proc. Natl. Acad. Sci. U.S.A. 83:4710-4714; Kimura et al., 1997, Science 277:942-946). An insulin receptor homolog has been characterized in *Drosophila*, termed DIR (*Drosophila* insulin receptor) (Ruan et al., 1995, J. Biol. Chem. 270:4236-4243), which exhibits extensive homology with vertebrate insulin and IGF receptors.

Recent discoveries from studies of *C. elegans* have also led to the identification of components involved in a presumptive insulin signaling pathway and have shown clear connections of this pathway to important aspects of metabolic regulation. (reviewed in Riddle and Albert, 1997, *C. elegans II*, Riddle et al., eds., Cold Spring Harbor Press, Plainview, New York, pp. 739-768). Molecular cloning has revealed that the *C. elegans daf-2* gene, is a nematode homolog of vertebrate insulin receptors. A *daf-2* mutant animal exhibits a dauer constitutive phenotype. The dauer stage is an alternative developmental stage that is induced when environmental factors are not adequate to promote successful reproduction in *C. elegans*. Dauer larvae remain relatively motionless, stop feeding, have increased deposition of fat, remain small in size, and are reproductively immature (O'Riordan and Burnell, 1989, Comp. Biochem. Physiol. 92B:233-238). Two other genes, *age-1* and *daf-16*, have been placed in the same pathway as *daf-2* based on analysis of genetic interactions (Morris et al., 1996, Nature 382:536-539; Ogg et al., 1997, Nature 389:994-999; Lin et al., 1997, Science 278:1319-1322). The *age-1* gene encodes a nematode homolog of PI3K, and the action of age-1 is required for the propagation of a daf-2 signal, in keeping with the role of PI3K in insulin signaling. Conversely, genetic analysis has shown that the normal role of daf-16 is one of blocking a signal generated by activated daf-2, and *daf-16* has been found to encode a homolog of the HNF-3/forkhead family of transcription factors.

There is another intriguing aspect to the phenotype of nematodes defective in components of the *daf-2* pathway with respect to effects on the life-span of the organism (normally about 14 days). Mutations in *daf-2* and *age-1* can more than double the life-span of animals, even under conditions that do not induce the formation of dauer larvae, and the

extension of life-span caused by *daf-2* or *age-1* mutations requires the activity of the *daf-16* gene (Lin et al., 1997, *Id.*; Tissenbaum and Ruvkun, 1998, Genetics 148:703-717; Larsen et al., 1995, Genetics 139:1567-1583).

Kawano et al., February 1, 1998, Worm Breeder's Gazette 15(2), 47, disclose the sequences of the A and B chain of two *C. elegans* insulin-like proteins. Ruvkun et al. disclose the nucleotide and protein sequences of several *C. elegans* insulin-like genes (Int'l Publication No. WO 98/51351, Int'l Publication Date November 19, 1998). Genbank® Accession Numbers (in parentheses) corresponding to: for ZK75.1 (AAC 46744 & GI 733563); ZK75.2 (AAC 46745 & GI 733561); ZK75.3 (AAC 46746 & GI 733562); ZK84.6 (AAC 48208 & GI 2914123); ZK1251.2 (CAA 92498 & GI 3881514); C17C3.4 (AAB 52688 & GI 1086914); MO4D8.2 (CAA 83611 & GI 3878561); MO4D8.3 (CAA 83609 & GI 3878559); F56F3.6 (CAA 83603 & GI 3877712); and T28B8.N (CAB 03444 & GI 3880317) disclose sequences that are not annotated as insulin-like genes. Citation of these references shall not be construed as an admission by applicant that they are available as prior art to the claimed invention.

SUMMARY OF THE INVENTION

The invention is directed to purified *C. elegans* insulin-like proteins, or derivatives or fragments thereof that display one or more functional activities of a *C. elegans* insulin-like proteins. The invention is also directed to compositions comprising such insulin-like protein or derivatives or fragments. The invention also concerns non-human animals comprising a transgene which encodes a *C. elegans* insulin-like protein. In preferred embodiments, the *C. elegans* insulin-like protein comprises an amino acid sequence selected from the group consisting of any one of SEQ ID NOs:1-18, 158-161, or 198-206.

The invention also directed to nucleic acids encoding *C. elegans* insulin-like proteins, such as a nucleic acid comprising a nucleotide sequence selected from the group consisting of any one of SEQ ID NOs:19-36, 162-165, and 207-215, or the complement thereof.

The invention also concerns methods of analyzing insulin expression or mis-expression comprising observing a nematode for the effects of expression or mis-expression of a *C. elegans* insulin-like protein, or derivative or fragment thereof that displays one or more functional activities of a *C. elegans* animal, wherein said *C. elegans* insulin-like protein has an amino acid sequence selected from the group consisting of any one of SEQ ID NOs:1-18, 158-161, or 198-206.

In preferred embodiments the *C. elegans* insulin-like protein is a member of Class IV.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1. Structural organization of precursor forms of the insulin superfamily of hormones are illustrated. The different domains that make up precursor forms of insulin-like hormones are represented as boxes labeled Pre, B, C, A, D, and E, extending from the N-terminus (left) to the C-terminus (right) of the nascent polypeptide chain, respectively. Domains that may remain in a mature hormone are represented as unshaded boxes (the B, A, and D peptide domains) or as lightly hatched (the C or "connecting" peptide domain). Domains that are removed during proteolytic processing are represented as shaded (the Pre peptide domain) or as hatched (the E peptide domain). IGF hormones are unique in having D and E peptide domains; these domains are represented as smaller boxes. Cleavage sites utilized by proteases during proteolytic processing (*i.e.*, protein maturation) are indicated below the boxes. The asterisk marks the position of cleavage by signal peptidase. Arrows indicate cleavage sites by prohormone convertases. Disulfide bonds (S-S) are represented above the boxes with lines indicating connections between covalently-bonded Cys residues.

FIG. 2. Conserved structural features of insulin superfamily members are shown, including aligned sequences of A and B peptide domains from diverse insulin superfamily. The alignment highlights the arrangement of conserved amino acid positions and their relationship to the overall folding pattern of the protein. The common helical regions found in the A and B chains are indicated by the symbol "<--->".

FIG. 3. Alignment of the *C. elegans* insulin-like protein family.

FIG. 4. Annotated sequence of *C. elegans* insulin-like protein F13B12.N and corresponding cDNA.

FIG. 5. Annotated sequence of *C. elegans* insulin-like protein ZK75.1 and corresponding cDNA.

FIG. 6. Annotated sequence of *C. elegans* insulin-like protein ZK75.2 and corresponding cDNA.

FIG. 7. Annotated sequence of *C. elegans* insulin-like protein ZK75.3 and corresponding cDNA.

FIG. 8. Annotated sequence of *C. elegans* insulin-like protein ZK84.6 and corresponding cDNA.

FIG. 9. Annotated sequence of *C. elegans* insulin-like protein ZK84.N2 and corresponding cDNA.

FIG. 10. Annotated sequence of *C. elegans* insulin-like protein ZK1251.2 and corresponding cDNA.

FIG. 11. Annotated sequence of *C. elegans* insulin-like protein ZK1251.N and corresponding cDNA.

FIG. 12. Annotated sequence of *C. elegans* insulin-like protein C06E2.N and corresponding cDNA.

FIG. 13. Annotated sequence of *C. elegans* insulin-like protein C17C3.4 and corresponding cDNA.

FIG. 14. Annotated sequence of *C. elegans* insulin-like protein C17C3.N and corresponding cDNA.

FIG. 15. Annotated sequence of *C. elegans* insulin-like protein M04D8.1 and corresponding cDNA.

FIG. 16. Annotated sequence of *C. elegans* insulin-like protein M04D8.2 and corresponding cDNA.

FIG. 17. Annotated sequence of *C. elegans* insulin-like protein M04D8.3 and corresponding cDNA.

FIG. 18. Annotated sequence of *C. elegans* insulin-like protein ZK84.N and corresponding cDNA.

FIG. 19. Annotated sequence of *C. elegans* insulin-like protein F56F3.6 and corresponding cDNA.

FIG. 20. Annotated sequence of *C. elegans* insulin-like protein T28B8.N and corresponding cDNA.

FIG. 21. Annotated sequence of *C. elegans* insulin-like protein ZC334.N and corresponding cDNA.

FIG. 22. Annotated sequence of *C. elegans* insulin-like protein TO8G5.N and corresponding cDNA.

FIG. 23. Annotated sequence of *C. elegans* insulin-like protein F41G3.N and corresponding cDNA.

FIG. 24. Annotated sequence of *C. elegans* insulin-like protein F41G3.N2 and corresponding cDNA.

FIG. 25. Annotated sequence of *C. elegans* insulin-like protein C17C3.N2 and corresponding cDNA.

FIG. 26. Annotated sequence of *C. elegans* insulin-like protein ZC334.N2 and corresponding cDNA.

FIG. 27. Annotated sequence of *C. elegans* insulin-like protein ZC334.N3 and corresponding cDNA.

FIG. 28. Annotated sequence of *C. elegans* insulin-like protein ZC334.N4 and corresponding cDNA.

FIG. 29. Annotated sequence of *C. elegans* insulin-like protein ZC334.N5 and corresponding cDNA.

FIG. 30. Annotated sequence of *C. elegans* insulin-like protein ZC334.N6 and corresponding cDNA.

FIG. 31. Annotated sequence of *C. elegans* insulin-like protein ZC334.N7 and corresponding cDNA.

FIG. 32A-32C. Annotated sequence of *C. elegans* insulin-like protein T10D4.N and corresponding cDNA.

FIG. 33. Annotated sequence of *C. elegans* insulin-like protein T10D4.N2 and corresponding cDNA.

FIG. 34. Annotated sequence of *C. elegans* insulin-like protein Y52A1.N and corresponding cDNA.

DETAILED DESCRIPTION OF THE INVENTION

In a desire to identify new and useful tools for probing the function and regulation of the insulin signaling pathway, an extensive search for insulin-like genes in the genome of *C. elegans* was conducted. The results of this search have revealed a surprisingly large and diverse family of insulin-like genes. These new insulin-like genes in *C. elegans* constitute very useful tools for probing the function and regulation of their corresponding pathways.

Systematic genetic analysis of signaling pathways involving insulin-like proteins in *C. elegans* can be expected to lead to the discovery of new drug targets, therapeutic proteins, diagnostics and prognostics useful in the treatment of diseases and clinical problems associated with the function of insulin superfamily hormones in humans and other animals, as well as clinical problems associated with aging and senescence. Furthermore, analysis of these same pathways using *C. elegans* insulin-like proteins as tools will have utility for identification and validation of pesticide targets in invertebrate pests that are components of these signaling pathways.

Use of *C. elegans* insulin-like genes for such purposes has advantages over manipulation of other known components of the nematode *daf-2* pathway, such as *daf-2*, *daf-16*, and *age-1*. Use of ligand-encoding *C. elegans* insulin-like genes will provide a superior approach for identifying factors that are upstream of the receptor in the signal

transduction pathway. Specifically, components involved in the synthesis, activation and turnover of insulin-like proteins may be identified. Furthermore, the large number of different insulin-like hormones could provide a means to separate components involved in response to different, specific environmental signals which may not be technically feasible with manipulation of downstream components of the pathway found in target tissues. Further, the diversity of different insulin-like hormones may provide a means to identify new receptor and/or signal transduction systems for insulin superfamily hormones that are structurally different from those that have been characterized to date in either vertebrates or invertebrates. Finally, use of *C. elegans* as a system for analyzing the function and regulation of insulin-like genes has great advantages over approaches in other organisms due to the ability to rapidly carry out large-scale, systematic genetic screens as well as the ability to screen small molecule libraries directly on whole organisms for possible therapeutic or pesticide use.

One advantage of investigating insulin-like genes in *C. elegans* comes from the tremendous progress made in the genome project for this organism. At the time of this writing, approximately 90% of the *C. elegans* genome has been sequenced, and that data is publically available in GenBank®, as well as in a specialized database for the *C. elegans* genome referred to as ACEDB (*i.e.*, A *C. elegans* Data Base) (Waterston and Sulston, 1995. "The genome of *Caenorhabditis elegans*", Proc. Natl. Acad. Sci. U.S.A. 92:10836-10840). In spite this wealth of genomic sequence information, the process of identifying authentic insulin superfamily genes in *C. elegans* is not trivial.

There are a number of factors that made identifying insulin-like genes in *C. elegans* genomic data particularly difficult. The insulin superfamily is fairly divergent at the sequence level and the degree of sequence homology between vertebrate and *C. elegans* insulin-like proteins is low. Furthermore, there are significant structural deviations in *C. elegans* insulin-like proteins that are absent or not common in the well-characterized vertebrate insulin-like proteins.

There are a number of software tools that can aid the process of identifying gene homologs in the *C. elegans* genome, including gene prediction programs (*e.g.*, GeneFinder), sequence homology searching programs (*e.g.*, BLAST, FASTA) and protein motif searching programs (*e.g.*, Prosite, BLOCKS, Markoff models). Nonetheless, identifying insulin-like genes within the *C. elegans* genome posed a significant challenge that went beyond just the straightforward application of any of these programs, due to the level of sequence divergence and structural variation. These problems were confounded further by the fact that insulins are small genes whose coding regions are often divided into smaller exons.

Small genes and exons are the most difficult to reliably predict from genomic sequence data with gene finding programs, and small blocks of divergent sequence are difficult to identify with homology searching programs as authentic sequence matches over those that would occur by chance.

The Prosite sequence matches found in the *C. elegans* genome illustrate the above-described problem. A pattern of specific amino acid residues has been derived from comparison of insulin superfamily proteins, termed an "insulin family signature," that reflects highly-conserved amino acid positions within the A chain of known insulin molecules. There are 27 matches to the Prosite "insulin family signature" identified in the *C. elegans* genome sequence and listed in ACEDB. Subsequent searches and analysis of insulin-like genes has revealed that only five of the 27 Prosite matches correspond to authentic insulin-like genes (as judged by criteria described below). Furthermore, at least another 17 authentic insulin-like genes in *C. elegans* did not have matches to the Prosite insulin family signature.

Given the difficulties in identifying insulin-like genes in the *C. elegans* genome, we pursued a strategy of combining several tools to find and evaluate potential insulin superfamily genes. Our search strategy used sequence features of known insulin superfamily genes, but focused initially on identifying matches to either: (1) B peptide region alone; (2) A peptide region alone; or (3) B and A peptide sequences fused together (*i.e.*, artificially). The A and B peptide regions (*i.e.*, domains) of known insulin superfamily proteins were chosen as queries since these are the most highly-conserved regions among the superfamily. The searching programs that were employed for the initial canvassing of the *C. elegans* genomic sequence included BLAST, FASTA, Markoff model searches, and exact pattern match searches (*i.e.*, regular expression searches). For matches to the B or A peptide alone, the genomic sequence was examined manually, and with the aid of the GeneFinder program, to identify a plausible nearby region encoding the other peptide in the correct relative position (*i.e.*, B peptide region N-terminal to A peptide region).

In most cases, the B and A peptide matches did not form a continuous open reading frame in the genomic DNA, and so the sequence was examined manually, and with the aid of a GeneFinder program, for the presence of likely splice junctions that would join the presumptive B and A peptide coding regions in-frame. Coding sequences N-terminal to presumptive B peptide coding regions were further examined manually, and with the aid of the GeneFinder program, for extended coding regions that might have a characteristic signal sequence for secretion following an initiator methionine (Met) codon. Also, regions upstream of the presumptive B peptide were examined manually, and with the aid of the

GeneFinder program, for potential splice sites that might join these segments to mRNA leaders found in trans-spliced mRNAs.

Each genomic match with correctly-oriented B and A peptides was further evaluated as follows to confirm that these regions preserved most of the structural features that are important for the formation of the characteristic insulin secondary and tertiary structure: (1) number and spacing of Cys residues involved in inter-chain and intra-chain disulfide bonds; (2) hydrophobic residues that form the "insulin core" at the interface of the A and B chains; (3) presence of Pro and Gly residues that promote characteristic breaks or turns between secondary structure elements; and (4) presence of proteolytic processing signals for maturation of the prehormone, especially removal of a C peptide, or regions preceding the B peptide and following a secretory signal.

This strategy resulted in the identification of at least 31 insulin-like genes. The structure and expression of the coding regions of 22 of these putative *C. elegans* insulin-like genes have been confirmed using an experimental approach involving reverse transcription of *C. elegans* mRNA, PCR amplification of specific cDNAs, cloning, and DNA sequencing. The details of the conditions used for each putative insulin-like gene are described in the Examples section below. Various non-limiting embodiments of the invention and applications and uses of these novel *C. elegans* insulin-like genes and proteins are described herein.

In a preferred embodiment, the invention provides a method of analyzing an effect of expression or mis-expression of a *C. elegans* insulin-like gene comprising observing a first nematode genetically engineered to express or mis-express a *C. elegans* insulin-like protein of any one of groups I, II or IV, or a derivative or fragment thereof that displays one or more functional activities of the *C. elegans* insulin-like protein. In another specific embodiment, the *C. elegans* protein is of group I.

In yet another specific embodiment, the claimed methods and products do not involve the proteins or nucleic acids of SEQ ID NOs: 6, 12, 24, or 30.

Isolation of *C. elegans* insulin-like genes

The invention relates to the nucleotide sequences of *C. elegans* insulin-like nucleic acids. In one embodiment, the insulin-like nucleic acids encode an insulin-like protein comprising the sequence of any one of SEQ ID NOs:1-18, 158-161, and 198-206. In another aspect, the invention provides a nucleic acid comprising a nucleotide sequence encoding at least a portion of an insulin-like protein, wherein the portion consists of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 60, or 100 continguous residues of any one of

SEQ ID NOs: 1-18, 158-161, and 198-206. In a more specific embodiment, the nucleotide sequences comprise at least 8 continguous nucleotides (*i.e.*, a hybridizable portion) of the cDNA sequences of any one of SEQ ID NOs:19-36, 162-165, and 207-215. In a preferred aspect, the nucleic acid sequences encode a Class IV *C. elegans* insulin-like polypeptide having the structure of a Class IV polypeptide (as further described in Example 2 below), such as the polypeptide defined by the amino acid sequence of any one of SEQ ID NOs:12-15, 18, or 198-203. Preferably, the nucleic acids consist of at least 10 (continguous) nucleotides, 25, nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, or 200 nucleotides or 300 nucleotids of an insulin-like sequence, or a full-length insulin-like coding sequence. In another embodiment, a nucleic acids comprising at least a portion of a *C. elegans* insulin-like nucleic acid of the invention is smaller than 100, 200, 500, 10,000, 15,000, 20,000 or 30,000 nucleotides in length. Nucleic acids can be single or double stranded. The invention also relates to nucleic acids hybridizable to or complementary to the foregoing sequences. In specific aspects, nucleic acids are provided which comprise a sequence complementary to at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of an insulin-like gene.

Hybridization conditions

In a specific embodiment, a nucleic acid which is hybridizable to an insulin-like nucleic acid (*e.g.*, having a sequence as set forth in SEQ ID NOs:19-36, 162-165, and 207-215), or to a nucleic acid encoding an insulin-like derivative, under conditions of low stringency is provided. By way of example and not limitation, procedures using such conditions of low stringency are as follows (*see also* Shilo and Weinberg, 1981, Proc. Natl. Acad. Sci. U.S.A. 78, 6789-6792). Filters containing DNA are pretreated for 6 h at 40°C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 µg/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X 10^6 cpm ^{32}P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at 40°C, and then washed for 1.5 h at 55°C in a solution containing 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 1.5 h at 60°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and re-exposed to film. Other conditions of low stringency which may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations).

In another specific embodiment, a nucleic acid which is hybridizable to an insulin-like nucleic acid under conditions of high stringency is provided. By way of example and not limitation, procedures using such conditions of high stringency are as follows. Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 μg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C in prehybridization mixture containing 100 μg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Washing of filters is done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is followed by a wash in 0.1X SSC at 50°C for 45 min before autoradiography. Other conditions of high stringency which may be used are well known in the art.

In another specific embodiment, a nucleic acid which is hybridizable to an insulin-like nucleic acid under conditions of moderate stringency is provided. Selection of appropriate conditions for such stringencies is well known in the art (*see e.g.*, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; *see also*, Ausubel et al., eds., in the Current Protocols in Molecular Biology series of laboratory technique manuals, © 1987-1997 Current Protocols, © 1994-1997 John Wiley and Sons, Inc.).

Nucleic acids encoding derivatives and analogs of insulin-like proteins, and insulin-like antisense nucleic acids are additionally provided. As is readily apparent, as used herein, a "nucleic acid encoding a fragment or portion of an insulin-like protein" shall be construed as referring to a nucleic acid encoding only the recited fragment or portion of the insulin-like protein and not the other contiguous portions of the insulin-like protein as a continuous sequence.

Fragments of insulin-like nucleic acids comprising regions conserved between (*i.e.*, with homology to) other insulin-like nucleic acids, of the same or different species, are also provided. Nucleic acids encoding one or more insulin-like protein domains are provided.

Cloning procedures

For expression cloning, an expression library can be constructed using known methods. For example, mRNA is isolated, cDNA is made and ligated into an expression vector (*e.g.*, a bacteriophage derivative) such that it is capable of being expressed by the host cell into which it is then introduced. Various screening assays can then be used to select for the expressed insulin-like product. In one embodiment, anti-insulin-like antibodies can be used for selection.

In another embodiment, polymerase chain reaction (PCR) is used to amplify the desired sequence in a genomic or cDNA library, prior to selection. Oligonucleotide primers representing known insulin-like sequences can be used as primers in PCR. In a preferred aspect, the oligonucleotide primers represent at least part of conserved segments of strong homology between insulin-like genes of different species. The synthetic oligonucleotides may be utilized as primers to amplify sequences from a source (RNA or DNA), preferably a cDNA library, of potential interest. PCR can be carried out, *e.g.*, by use of a Perkin-Elmer Cetus thermal cycler and Taq polymerase (*e.g.*, Gene Amp™). The nucleic acid being amplified can include mRNA or cDNA or genomic DNA from any species. One may synthesize degenerate primers for amplifying homologs from other species in the PCR reactions. It is also possible to vary the stringency of hybridization conditions used in priming the PCR reactions, to allow for greater or lesser degrees of nucleotide sequence similarity between the known insulin-like nucleotide sequences and a nucleic acid homolog (or ortholog) being isolated. For cross species hybridization, low stringency conditions are preferred. For same species hybridization, moderately stringent conditions are preferred. After successful amplification of a segment of an insulin-like homolog, that segment may be cloned and sequenced by standard techniques, and utilized as a probe to isolate a complete cDNA or genomic clone. This, in turn, permits the determination of the gene's complete nucleotide sequence, the analysis of its expression, and the production of its protein product for functional analysis, as described below. In this fashion, additional genes encoding insulin-like proteins and insulin-like analogs may be identified.

The above-described methods are not meant to limit the following general description of methods by which clones of insulin-like genes may be obtained.

Any eukaryotic cell potentially can serve as the nucleic acid source for molecular cloning of an insulin-like gene. The nucleic acid sequences encoding insulin-like proteins may be isolated from vertebrate, mammalian, human, porcine, bovine, feline, avian, equine, canine, as well as additional primate sources, insects (*e.g., Drosophila*), invertebrates (*e.g., C. elegans*), plants, *etc.* The DNA may be obtained by standard procedures known in the art from cloned DNA (*e.g.*, a DNA "library"), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from the desired cell (*see e.g.*, Sambrook et al., supra; Glover (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II.) Clones derived from genomic DNA may contain regulatory and intron DNA regions in addition to coding regions; clones derived from cDNA will contain only exon sequences. Whatever the source, the gene should be molecularly cloned into a suitable vector for propagation of the gene.

In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNAse in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques, such as agarose and polyacrylamide gel electrophoresis and column chromatography.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene may be accomplished in a number of ways. For example, if a portion of an insulin-like gene or its specific RNA or a fragment thereof is available and can be purified and labeled, the generated DNA fragments may be screened by nucleic acid hybridization to the labeled probe (e.g. Benton and Davis, 1977, Science 196:180). Those DNA fragments with substantial homology to the probe will hybridize. It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of fragment sizes with those expected according to a known restriction map if such is available. Further selection can be carried out on the basis of the properties of the gene. Alternatively, the presence of the desired gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper mRNAs, can be selected and expressed to produce a protein that has, *e.g.*, similar or identical electrophoretic migration, isoelectric focusing behavior, proteolytic digestion maps, hormonal activity, binding activity, or antigenic properties as known for an insulin-like protein. Using an antibody to a known insulin-like protein, other insulin-like proteins may be identified by binding of the labeled antibody to expressed putative insulin-like proteins, *e.g.*, in an ELISA (enzyme-linked immunosorbent assay)-type procedure. Further, using a binding protein specific to a known insulin-like protein, other insulin-like proteins may be identified by binding to such a protein (*see e.g.*, Clemmons, 1993, Mol. Reprod. Dev. 35:368-374; Loddick et al., 1998, Proc. Natl. Acad. Sci. U.S.A. 95:1894-1898).

An insulin-like gene can also be identified by mRNA selection using nucleic acid hybridization followed by *in vitro* translation. In this procedure, fragments are used to isolate complementary mRNAs by hybridization. Such DNA fragments may represent available, purified insulin-like DNA of another species (*e.g.*, *Drosophila*, mouse, human). Immunoprecipitation analysis or functional assays (*e.g.*, aggregation ability *in vitro*, binding to receptor, *etc.*) of the *in vitro* translation products of the isolated products of the isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments that

contain the desired sequences. In addition, specific mRNAs may be selected by adsorption of polysomes isolated from cells to immobilized antibodies specifically directed against insulin-like protein. A radiolabeled insulin-like cDNA can be synthesized using the selected mRNA (from the adsorbed polysomes) as a template. The radiolabeled mRNA or cDNA may then be used as a probe to identify the insulin-like DNA fragments from among other genomic DNA fragments.

Alternatives to isolating the insulin-like genomic DNA include, chemically synthesizing the gene sequence itself from a known sequence or making cDNA to the mRNA which encodes the insulin-like protein. For example, RNA for cDNA cloning of the insulin-like gene can be isolated from cells which express the gene.

The identified and isolated gene can then be inserted into an appropriate cloning vector. A large number of vector-host systems known in the art may be used. Possible vectors include plasmids or modified viruses, but the vector system must be compatible with the host cell used. Such vectors include bacteriophages such as lambda derivatives, or plasmids such as PBR322 or pUC plasmid derivatives or the Bluescript vector (Stratagene). The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and an insulin-like gene may be modified by homopolymeric tailing. Recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, *etc.*, so that many copies of the gene sequence are generated.

In an alternative method, the desired gene may be identified and isolated after insertion into a suitable cloning vector in a "shot gun" approach. Enrichment for the desired gene, for example, by size fractionization, can be done before insertion into the cloning vector.

In specific embodiments, transformation of host cells with recombinant DNA molecules that incorporate an isolated insulin-like gene, cDNA, or synthesized DNA sequence enables generation of multiple copies of the gene. Thus, the gene may be obtained in large quantities by growing transformants, isolating the recombinant DNA molecules from the transformants and, when necessary, retrieving the inserted gene from the isolated recombinant DNA.

The insulin-like sequences provided by the instant invention include those nucleotide sequences encoding substantially the same amino acid sequences as found in native insulin-like proteins, and those encoded amino acid sequences with functionally equivalent amino acids, as well as those encoding other insulin-like derivatives or analogs, as described below for insulin-like derivatives and analogs.

Expression of *C. elegans* insulin-like genes

The nucleotide sequence coding for an insulin-like protein or a functionally active analog or fragment or other derivative thereof, can be inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. The necessary transcriptional and translational signals can also be supplied by the native insulin-like gene and/or its flanking regions. A variety of host-vector systems may be utilized to express the protein-coding sequence such as mammalian cell systems infected with virus (*e.g.*, vaccinia virus, adenovirus, *etc.*); insect cell systems infected with virus (*e.g.*, baculovirus); microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their strengths and specificities. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation elements may be used. In yet another embodiment, a fragment of an insulin-like protein comprising one or more domains of the insulin-like protein is expressed.

Any of the methods previously described for the insertion of DNA fragments into a vector may be used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional/translational control signals and the protein coding sequences. These methods may include *in vitro* recombinant DNA and synthetic techniques and *in vivo* recombinants (genetic recombination). Expression of a nucleic acid sequence encoding an insulin-like protein or peptide fragment may be regulated by a second nucleic acid sequence so that the insulin-like protein or peptide is expressed in a host transformed with the recombinant DNA molecule. For example, expression of an insulin-like protein may be controlled by any promoter/enhancer element known in the art. Promoters which may be used to control insulin-like gene expression include the SV40 early promoter region, the promoter contained in the 3' long terminal repeat of Rous sarcoma, the herpes thymidine kinase promoter, the regulatory sequences of the metallothionein gene; prokaryotic expression vectors such as the β-lactamase promoter, or the *lac* promoter; plant expression vectors comprising the nopaline synthetase promoter or the cauliflower mosaic virus 35S

RNA promoter, and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase; promoter elements from yeast or other fungi such as the Gal 4 promoter, the alcohol dehydrogenase promoter, phosphoglycerol kinase promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., 1984, Cell 38:639-646); a gene control region which is active in pancreatic beta cells (Hanahan, 1985, Nature 315:115-122), an immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., 1984, Cell 38:647-658), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., 1986, Cell 45:485-495), albumin gene control region which is active in liver (Pinkert et al., 1987, Genes and Devel. 1:268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., 1985, Mol. Cell. Biol. 5:1639-1648); alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, Genes and Devel. 1:161-171), beta-globin gene control region which is active in myeloid cells (Mogram et al., 1985, Nature 315:338-340); myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., 1987, Cell 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, Nature 314:283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., 1986, Science 234:1372-1378).

In a specific embodiment, a vector is used that comprises a promoter operably linked to an insulin-like gene nucleic acid, one or more origins of replication, and, optionally, one or more selectable markers (*e.g.*, an antibiotic resistance gene).

Expression constructs can be made by subcloning an insulin-like coding sequence into the EcoRI restriction site of each of the three pGEX vectors (Smith and Johnson, 1988, Gene 7:31-40). This allows for the expression of the insulin-like protein product from the subclone in the correct reading frame.

Expression vectors containing insulin-like gene inserts can be identified by three general approaches: (a) nucleic acid hybridization; (b) presence or absence of "marker" gene functions; and (c) expression of inserted sequences. In the first approach, the presence of an insulin-like gene inserted in an expression vector can be detected by nucleic acid hybridization using probes comprising sequences that are homologous to an inserted insulin-like gene. In the second approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine kinase activity, resistance to antibiotics, transformation

phenotype, occlusion body formation in baculovirus, *etc.*) caused by the insertion of an insulin-like gene in the vector. For example, if the insulin-like gene is inserted within the marker gene sequence of the vector, recombinants containing the insulin-like insert can be identified by the absence of the marker gene function. In the third approach, recombinant expression vectors can be identified by assaying the insulin-like product expressed by the recombinant. Such assays can be based, for example, on the physical or functional properties of the insulin-like protein in *in vitro* assay systems, *e.g.*, binding with anti-insulin-like protein antibody.

Once a particular recombinant DNA molecule is identified and isolated, several methods known in the art may be used to propagate it. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. Some of the expression vectors which can be used include human or animal viruses such as vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (*e.g.*, lambda phage), and plasmid and cosmid DNA vectors.

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered insulin-like protein may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (*e.g.*, glycosylation, phosphorylation of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce a non-glycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in mammalian cells can be used to ensure "native" glycosylation of a heterologous protein. Furthermore, different vector/host expression systems may effect processing reactions to different extents.

In other embodiments of the invention, the insulin-like protein, fragment, analog, or derivative may be expressed as a fusion, or chimeric protein product (comprising the protein, fragment, analog, or derivative joined via a peptide bond to a heterologous protein sequence of a different protein). Such a chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric

product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, *e.g.*, by use of a peptide synthesizer.

Identification and purification of gene products

The invention provides compositions comprising amino acid sequences of insulin-like proteins and fragments and derivatives thereof which comprise an antigenic determinant (*i.e.*, can be recognized by an antibody) or which are otherwise functionally active, as well as nucleic acid sequences encoding the foregoing. "Functionally active" insulin-like material as used herein refers to that material displaying one or more functional activities associated with a full-length (wild-type) insulin-like protein, *e.g.*, binding to an insulin-like receptor (*e.g.*, *daf-2*) or insulin-like protein binding partner, antigenicity (binding to an anti-insulin-like protein antibody), immunogenicity, *etc.* The compositions may consist essentially of the insulin-like proteins and fragments and derivatives thereof. Alternatively, the insulin-like proteins and fragments and derivatives thereof may be a component of a composition that comprises other components, for example, a diluent such as saline, a pharmaceutically acceptable carrier or excipient, a culture medium, etc.

In specific embodiments, the invention provides fragments of an insulin-like protein consisting of at least 6 amino acids, 10 amino acids, 20 amino acids, 50 amino acids, or of at least 75 amino acids. In other embodiments, the proteins comprise or consist essentially of an insulin-like B peptide domain, an insulin-like A peptide domain, an insulin-like C peptide domain, or any combination of the foregoing, of an insulin-like protein. Fragments, or proteins comprising fragments, lacking some or all of the foregoing regions of a insulin-like protein are also provided. Nucleic acids encoding the foregoing are provided.

Once a recombinant which expresses the insulin-like gene sequence is identified, the gene product can be analyzed. This is achieved by assays based on the physical or functional properties of the product, including radioactive labeling of the product followed by analysis by gel electrophoresis, immunoassay, *etc.* The gene product may be isolated and purified by standard methods including chromatography (*e.g.*, ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. The functional properties may be evaluated using any suitable assay. The amino acid sequence of the protein can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant. As a result, the protein can be synthesized by standard chemical methods known in the art (*e.g.*, *see* Hunkapiller et al., 1984, Nature 310:105-111).

In an alternate embodiment, native insulin-like proteins can be purified from natural sources, by standard methods such as those described above (*e.g.*, immunoaffinity purification).

Insulin-like proteins, whether produced by recombinant DNA techniques or by chemical synthetic methods or by purification of native proteins, can include all or part of the amino acid sequence substantially as depicted in any of FIGs 4-36 (SEQ ID NOs:1-18, 158-161, and 198-206), as well as fragments and other derivatives, and analogs thereof, including proteins homologous thereto.

Structure of insulin-like genes and proteins

The structure of insulin-like genes and proteins of the invention can be analyzed by various methods known in the art, including genetic analysis and protein analysis.

Genetic analysis methods for determining the structure of cloned DNA or cDNA corresponding to an insulin-like include Southern hybridization, Northern hybridization, restriction endonuclease mapping, and DNA sequence analysis. Accordingly, this invention provides nucleic acid probes recognizing an insulin-like gene. For example, polymerase chain reaction followed by Southern hybridization with an insulin-like gene-specific probe can allow the detection of an insulin-like gene in DNA from various cell types. Methods of amplification other than PCR are commonly known and can also be employed. In one embodiment, Southern hybridization can be used to determine the genetic linkage of an insulin-like gene. Northern hybridization analysis can be used to determine the expression of an insulin-like gene. Various cell types, at various states of development or activity can be tested for insulin-like gene expression. The stringency of the hybridization conditions for both Southern and Northern hybridization can be manipulated to ensure detection of nucleic acids with the desired degree of relatedness to the specific insulin-like gene probe used. Modifications of these methods and other methods commonly known in the art can be used.

Restriction endonuclease mapping can be used to roughly determine the genetic structure of an insulin-like gene. Restriction maps derived by restriction endonuclease cleavage can be confirmed by DNA sequence analysis.

DNA sequence analysis can be performed by any techniques known in the art, such as the method of Maxam and Gilbert (1980, Meth. Enzymol. 65:499-560), the Sanger dideoxy method (Sanger et al., 1977, Proc. Natl. Acad. Sci. U.S.A. 74:5463), the use of T7 DNA polymerase (Tabor and Richardson, U.S. Patent No. 4,795,699), or use of an automated DNA sequenator (*e.g.*, Applied Biosystems, Foster City, California).

The amino acid sequence of an insulin-like protein can be derived by deduction from the DNA sequence, or alternatively, by direct sequencing of the protein, *e.g.*, with an automated amino acid sequencer. An insulin-like protein sequence can be further characterized by a hydrophilicity analysis (Hopp and Woods, 1981, Proc. Natl. Acad. Sci. U.S.A. 78:3824). A hydrophilicity profile can be used to identify the hydrophobic and hydrophilic regions of the insulin-like protein and the corresponding regions of the gene sequence which encode such regions.

Secondary, structural analysis (Chou and Fasman, 1974, Biochemistry 13:222) can also be done, to identify regions of an insulin-like protein that assume specific secondary structures.

Manipulation, translation, and secondary structure prediction, open reading frame prediction and plotting, as well as determination of sequence homologies, can also be accomplished using computer software programs available in the art.

Other methods of structural analysis include X-ray crystallography, nuclear magnetic resonance spectroscopy and computer modeling.

Antibodies to insulin-like protein

Insulin-like protein or its fragments (e.g. an insulin-like protein encoded by a sequence of any of SEQ ID NOs:1-18, 158-161, and 198-206, or a subsequence thereof), or other derivatives, or analogs thereof, may be used as an immunogen to generate antibodies. Such antibodies include polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library. In another embodiment, antibodies to a domain (*e.g.*, an insulin-like receptor binding domain) of an insulin-like protein are produced. In a specific embodiment, fragments of an insulin-like protein identified as hydrophilic are used as immunogens for antibody production using art-known methods. Some examples of suitable techniques include methods which provides for the production of antibody molecules by continuous cell lines in culture; the production of monoclonal antibodies in germ-free animals (*see e.g.*, PCT/US90/02545); the use of human hybridomas (Cole et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030); transforming human B cells with EBV virus *in vitro* (Cole et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, pp. 77-96). Additionally, known techniques can be used for the production of "chimeric antibodies" (e.g. by splicing the genes from a mouse antibody molecule specific for an insulin-like protein together with genes from a human antibody molecule of appropriate biological activity), insulin-like-specific single chain antibodies; and Fab expression libraries (e.g. to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity

for insulin-like proteins, derivatives, or analogs). The foregoing antibodies can be used against the insulin-like protein sequences described herein, *e.g.*, for imaging these proteins, measuring levels thereof, in diagnostic methods, *etc.*

Insulin-like proteins, derivatives and analogs

The invention further relates to insulin-like proteins and derivatives, fragments and analogs thereof which can be encoded by the nucleic acids described above. The insulin-like proteins comprise the amino acid sequence of any one of SEQ ID NOs 1-18, 158-161, and 198-206. In another aspect, the invention provides a protein consisting of or comprising at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, or 30 amino acid residues of any one of SEQ ID NOs: 1-18, 158-161, and 198-206. In a preferred aspect, the *C. elegans* insulin-like polypeptide has the structure of a Class IV polypeptide (as further described in Example 2 below), such as the polypeptide defined by the amino acid sequence of any one of SEQ ID NOs:12-15, 18, or 198-203. In particular aspects, the proteins, derivatives, or analogs are of insulin-like proteins of animals, *e.g.*, fly, frog, mouse, rat, pig, cow, dog, monkey, human, worm, or plant.

In a specific embodiment, the derivative or analog is functionally active, *i.e.*, capable of exhibiting one or more functional activities associated with a full-length, wild-type insulin-like protein. As one example, such derivatives or analogs which have the desired immunogenicity or antigenicity can be used in immunoassays, for immunization, for inhibition of insulin-like activity, *etc.* As another example, such derivatives or analogs which have the desired binding activity can be used for binding to the *daf-2* gene product. As yet another example, such derivatives or analogs which have the desired binding activity can be used for binding to a binding protein specific for a known insulin-like protein (*see e.g.*, Clemmons, 1993, Mol. Reprod. Dev. 35:368-374; Loddick et al., 1998, Proc. Natl. Acad. Sci. U.S.A. 95:1894-1898). Derivatives or analogs that retain, or alternatively lack or inhibit, a desired insulin-like protein property-of-interest (*e.g.*, binding to an insulin-like protein binding partner), can be used as inducers, or inhibitors, respectively, of such property and its physiological correlates. A specific embodiment relates to an insulin-like protein fragment that can be bound by an anti-insulin-like protein antibody. Derivatives or analogs of an insulin-like protein can be tested for the desired activity by procedures discussed herein and also those known in the art.

Insulin-like derivatives can be made by altering insulin-like sequences by substitutions, additions (*e.g.*, insertions) or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences, other DNA sequences

which encode substantially the same amino acid sequence as an insulin-like gene may be used in the practice of the present invention. These can include nucleotide sequences comprising all or portions of an insulin-like gene which is altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change. Likewise, the insulin-like derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of an insulin-like protein including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutions for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Such substitutions are generally understood to be conservative substitutions.

The invention also provides proteins consisting of or comprising a fragment of an insulin-like protein consisting of at least 6 (continguous) amino acids of the insulin-like protein. In other embodiments, the fragment consists of at least 10, at least 15, at least 20 or at least 50 amino acids of the insulin-like protein. In specific embodiments, such fragments are not larger than 35, 100 or 200 amino acids. Derivatives or analogs of insulin-like proteins include those molecules comprising regions that are substantially homologous to an insulin-like protein or fragment thereof (*e.g.*, in various embodiments, at least 60% or 70% or 80% or 90% or 95% identity over an amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art) or whose encoding nucleic acid is capable of hybridizing to a coding insulin-like gene sequence, under high stringency, moderate stringency, or low stringency conditions.

The insulin-like derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, a cloned insulin-like gene sequence can be modified by any of numerous strategies known in the art. The sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired,

isolated, and ligated *in vitro*. In the production of a modified gene encoding a derivative or analog of an insulin-like protein, care should be taken to ensure that the modified gene remains within the same translational reading frame as the native protein, uninterrupted by translational stop signals, in the gene region where the desired insulin-like protein activity is encoded.

Additionally, an insulin-like nucleic acid sequence can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or to form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used, including but not limited to, chemical mutagenesis, *in vitro* site-directed mutagenesis, use of TAB® linkers (Pharmacia), *etc.*

Manipulations of an insulin-like protein sequence may also be made at the protein level. Included within the scope of the invention are insulin-like protein fragments or other derivatives or analogs which are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, *etc.* Any of numerous chemical modifications may be carried out by known techniques, including but not limited to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, $NaBH_4$, acetylation, formylation, oxidation, reduction, metabolic synthesis in the presence of tunicamycin, *etc.*

In addition, analogs and derivatives of an insulin-like protein can be chemically synthesized. For example, a peptide corresponding to a portion of an insulin-like protein which comprises the desired domain, or which mediates the desired activity *in vitro*, can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the insulin-like sequence. Non-classical amino acids include the D-isomers of the common amino acids, α-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, γ-Abu, ϵ-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β-alanine, fluoro-amino acids, designer amino acids such as β-methyl amino acids, Cα-methyl amino acids, Nα-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

Chimeric or fusion proteins can be made comprising an insulin-like protein or fragment thereof (preferably consisting of at least a domain or motif of the insulin-like

protein, or at least 6, and preferably at least 10 amino acids of the insulin-like protein) joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein. Such a chimeric protein can be produced by any known method, including: recombinant expression of a nucleic acid encoding the protein (comprising an insulin-like-coding sequence joined in-frame to a coding sequence for a different protein); ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other in the proper coding frame, and expressing the chimeric product; and protein synthetic techniques, *e.g.*, by use of a peptide synthesizer.

The insulin-like derivative can be a molecule comprising a region of homology with a insulin-like protein. For example, a first protein region can be considered "homologous" to a second protein region when the amino acid sequence of the first region is at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, or 95% identical, when compared to any sequence in the second region of an equal number of amino acids as the number contained in the first region or when compared to an aligned sequence of the second region that has been aligned by a computer homology program known in the art. For example, a molecule can comprise one or more regions homologous to an insulin-like domain or a portion thereof.

A fragment of an insulin-like protein can be those fragments in the respective insulin-like proteins of the invention most homologous to specific fragments of a human or mouse insulin-like protein as identified by protein analysis methods.

Insulin-like fragments and derivatives of such fragments, may comprise or consist of one or more domains of an insulin-like protein, such as an insulin-like B peptide domain, an insulin-like A peptide domain, and/or an insulin-like connecting (C) peptide domain (or functional portion thereof). In particular examples, the insulin-like protein derivatives has either an A peptide domain or a B peptide domain. Such a protein may retain such domains separated by a peptide spacer. The spacer may be the same as or different from an insulin-like connecting (C) peptide.

A insulin-like protein derivative may comprises one or more domains (or functional portion(s) thereof) of an insulin-like protein, and a one or more mutant domains(*e.g.*, due to deletion or point mutation(s)) of an insulin-like protein (*e.g.*, such that the mutant domain has decreased function).

Proteins which interact with insulin-like proteins

The present invention further provides methods of identifying or screening for proteins which interact with *C. elegans* insulin-like proteins, or derivatives, fragments or analogs thereof. A preferred method is a yeast two hybrid assay system or a variation thereof. The yeast two-hybrid method has been used to analyze IGF-1-receptor interactions (*see* Zhu and Kahn, 1997, Proc. Natl. Acad. Sci. U.S.A. 94, 13063-13068). Derivatives (*e.g.*, fragments) and analogs of a protein can also be assayed for binding to a binding partner by any method known in the art, for example, immunoprecipitation with an antibody that binds to the protein in a complex followed by analysis by size fractionation of the immunoprecipitated proteins (*e.g.*, by denaturing polyacrylamide gel electrophoresis), Western analysis, non-denaturing gel electrophoresis, *etc.*

Known methods can be used for assaying and screening fragments, derivatives and analogs of *C. elegans* insulin-like protein interacting proteins (for binding to a *C. elegans* insulin-like peptide). Derivatives, analogs and fragments of proteins that interact with a *C. elegans* insulin-like protein can be identified by means of a yeast two hybrid assay system (Fields and Song, 1989, Nature 340:245-246 and U.S. Patent No. 5,283,173). Because the interactions are screened for in yeast, the intermolecular protein interactions detected in this system occur under physiological conditions that mimic the conditions in mammalian cells. This feature facilitates identification of proteins capable of interaction with a *C. elegans* insulin-like protein from species other than *C. elegans*.

Identification of interacting proteins by the improved yeast two hybrid system is based upon the detection of expression of a reporter gene, the transcription of which is dependent upon the reconstitution of a transcriptional regulator by the interaction of two proteins, each fused to one half of the transcriptional regulator. The "bait" (*i.e.*, *C. elegans* insulin-like protein or derivative or analog thereof) and "prey" proteins (proteins to be tested for ability to interact with the bait) are expressed as fusion proteins to a DNA binding domain, and to a transcriptional regulatory domain, respectively, or *vice versa*. In various specific embodiments, the prey has a complexity of at least about 50, about 100, about 500, about 1,000, about 5,000, about 10,000, or about 50,000; or has a complexity in the range of about 25 to about 100,000, about 100 to about 100,000, about 50,000 to about 100,000, or about 100,000 to about 500,000. For example, the prey population can be one or more nucleic acids encoding mutants of a protein (*e.g.*, as generated by site-directed mutagenesis or another method of making mutations in a nucleotide sequence). Preferably, the prey populations are proteins encoded by DNA, *e.g.*, cDNA or genomic DNA or synthetically-generated DNA. For example, the populations can be expressed from chimeric genes

comprising cDNA sequences from an un-characterized sample of a population of cDNA from mRNA. In one embodiment, recombinant biological libraries expressing random peptides can be used as the source of prey nucleic acids.

The invention provides methods of screening for inhibitors or enhancers of the protein interactants identified herein. Briefly, the protein-protein interaction assay can be carried out as described herein, except that it is done in the presence of one or more candidate molecules. An increase or decrease in reporter gene activity relative to that present when the one or more candidate molecules are absent indicates that the candidate molecule has an effect on the interacting pair. In a preferred method, inhibition of the interaction is selected for (*i.e.*, inhibition of the interaction is necessary for the cells to survive), for example, where the interaction activates the *URA3* gene, causing yeast to die in medium containing the chemical 5-fluoroorotic acid (Rothstein, 1983, Meth. Enzymol. 101:167-180). The identification of inhibitors of such interactions can also be accomplished, for example, using competitive inhibitor assays, as described above.

In general, proteins of the bait and prey populations are provided as fusion (chimeric) proteins (preferably by recombinant expression of a chimeric coding sequence) comprising each protein contiguous to a pre-selected sequence. For one population, the pre-selected sequence is a DNA binding domain. The DNA binding domain can be any DNA binding domain, as long as it specifically recognizes a DNA sequence within a promoter. For example, the DNA binding domain is of a transcriptional activator or inhibitor. For the other population, the pre-selected sequence is an activator or inhibitor domain of a transcriptional activator or inhibitor, respectively. The regulatory domain alone (i.e. not as a fusion to a protein sequence) and the DNA-binding domain alone preferably do not detectably interact (so as to avoid false positives in the assay). The assay system further includes a reporter gene operably linked to a promoter that contains a binding site for the DNA binding domain of the transcriptional activator (or inhibitor). Accordingly, in the present method of the present invention, binding of a *C. elegans* insulin-like fusion protein to a prey fusion protein leads to reconstitution of a transcriptional activator (or inhibitor) which activates (or inhibits) expression of the reporter gene. The activation (or inhibition) of transcription of the reporter gene occurs intracellularly, *e.g.*, in prokaryotic or eukaryotic cells, preferably in cell culture.

The promoter that is operably linked to the reporter gene nucleotide sequence can be a native or non-native promoter of the nucleotide sequence, and the DNA binding site(s) that are recognized by the DNA binding domain portion of the fusion protein can be native to the promoter (if the promoter normally contains such binding site(s)) or non-native to the

promoter. Thus, for example, one or more tandem copies (*e.g.*, four or five copies) of the appropriate DNA binding site can be introduced upstream of the TATA box in the desired promoter (*e.g.*, in the area of about position -100 to about -400). In a preferred aspect, 4 or 5 tandem copies of the 17 bp UAS (GAL4 DNA binding site) are introduced upstream of the TATA box in the desired promoter, which is upstream of the desired coding sequence for a selectable or detectable marker. In a preferred embodiment, the GALl-10 promoter is operably fused to the desired nucleotide sequence; the GALl-10 promoter already contains 5 binding sites for GAL4.

Alternatively, the transcriptional activation binding site of the desired gene(s) can be deleted and replaced with GAL4 binding sites (Bartel et al., 1993, BioTechniques 14:920-924, Chasman et al., 1989, Mol. Cell. Biol. 9:4746-4749). The reporter gene preferably contains the sequence encoding a detectable or selectable marker, the expression of which is regulated by the transcriptional activator, such that the marker is either turned on or off in the cell in response to the presence of a specific interaction. Preferably, the assay is carried out in the absence of background levels of the transcriptional activator (*e.g.*, in a cell that is mutant or otherwise lacking in the transcriptional activator). More than one reporter gene can be used to detect transcriptional activation, *e.g.*, one reporter gene encoding a detectable marker and one or more reporter genes encoding different selectable markers. The detectable marker can be any molecule that can give rise to a detectable signal, *e.g.*, a fluorescent protein or a protein that can be readily visualized or that is recognizable by a specific antibody. The selectable marker can be any protein molecule that confers the ability to grow under conditions that do not support the growth of cells not expressing the selectable marker, *e.g.*, the selectable marker is an enzyme that provides an essential nutrient and the cell in which the interaction assay occurs is deficient in the enzyme and the selection medium lacks such nutrient. The reporter gene can either be under the control of the native promoter that naturally contains a binding site for the DNA binding protein, or under the control of a heterologous or synthetic promoter.

The activation domain and DNA binding domain used in the assay can be from a wide variety of transcriptional activator proteins, as long as these transcriptional activators have separable binding and transcriptional activation domains. For example, the GAL4 protein of *S. cerevisiae* (Ma et al., 1987, Cell 48:847-853), the GCN4 protein of *S. cerevisiae* (Hope and Struhl, 1986, Cell 46:885-894), the ARD1 protein of *S. cerevisiae* (Thukral et al., 1989, Mol. Cell. Biol. 9:2360-2369), and the human estrogen receptor (Kumar et al., 1987, Cell 51:941-951), have separable DNA binding and activation domains. The DNA binding domain and activation domain that are employed in the fusion

proteins need not be from the same transcriptional activator. In a specific embodiment, a GAL4 or LEXA DNA binding domain is employed. In another specific embodiment, a GAL4 or herpes simplex virus VP16 (Triezenberg et al., 1988, Genes Dev. 2:730-742) activation domain is employed. In a specific embodiment, amino acids 1-147 of GAL4 (Ma et al., 1987, Cell 48:847-853; Ptashne et al., 1990, Nature 346:329-331) is the DNA binding domain, and amino acids 411-455 of VP16 (Triezenberg et al., 1988, Genes Dev. 2:730-742; Cress et al., 1991, Science 251:87-90) comprise the activation domain.

In a preferred embodiment, the yeast transcription factor GAL4 is reconstituted by protein-protein interaction and the host strain is mutant for GAL4. In another embodiment, the DNA-binding domain is Ace1N and/or the activation domain is Ace1, the DNA binding and activation domains of the Ace1 protein, respectively. Ace1 is a yeast protein that activates transcription from the *CUP1* operon in the presence of divalent copper. *CUP1* encodes metallothionein, which chelates copper, and the expression of CUP1 protein allows growth in the presence of copper, which is otherwise toxic to the host cells. The reporter gene can also be a *CUP1-lacZ* fusion that expresses the enzyme beta-galactosidase (detectable by routine chromogenic assay) upon binding of a reconstituted Ace1N transcriptional activator (*see* Chaudhuri et al., 1995, FEBS Letters 357:221-226). In another embodiment, the DNA binding domain of the human estrogen receptor is used, with a reporter gene driven by one or three estrogen receptor response elements (Le Douarin et al., 1995, Nucl. Acids. Res. 23:876-878).

The DNA binding domain and the transcriptional activator/inhibitor domain each preferably has a nuclear localization signal (*see* Ylikomi et al., 1992, EMBO J. 11:3681-3694, Dingwall and Laskey, 1991, TIBS 16:479-481) functional in the cell in which the fusion proteins are to be expressed.

To facilitate isolation of the encoded proteins, the fusion constructs can further contain sequences encoding affinity tags such as glutathione-S-transferase or maltose-binding protein or an epitope of an available antibody, for affinity purification (*e.g.*, binding to glutathione, maltose, or a particular antibody specific for the epitope, respectively) (Allen et al., 1995, TIBS 20:511-516). In another embodiment, the fusion constructs further comprise bacterial promoter sequences for recombinant production of the fusion protein in bacterial cells.

The host cell in which the interaction assay occurs can be any cell, prokaryotic or eukaryotic, in which transcription of the reporter gene can occur and be detected such as mammalian (*e.g.*, monkey, mouse, rat, human, bovine), chicken, bacterial, or insect cells, and is preferably a yeast cell. Expression constructs encoding and capable of expressing the

binding domain fusion proteins, the transcriptional activation domain fusion proteins, and the reporter gene product(s) are provided within the host cell, by mating of cells containing the expression constructs, or by cell fusion, transformation, electroporation, microinjection, *etc*. When the assay is carried out in mammalian cells (*e.g.*, hamster cells, HeLa cells), the DNA binding domain can be the GAL4 DNA binding domain, the activation domain can be the herpes simplex virus VP16 transcriptional activation domain, and the reporter gene can contain the desired coding sequence operably linked to a minimal promoter element from the adenovirus E1B gene driven by several GAL4 DNA binding sites (*see* Fearon et al., 1992, Proc. Natl. Acad. Sci. U.S.A. 89:7958-7962). The host cell used should not express an endogenous transcription factor that binds to the same DNA site as that recognized by the DNA binding domain fusion population. Also, preferably, the host cell is mutant or otherwise lacking in an endogenous, functional form of the reporter gene(s) used in the assay.

Various vectors and host strains for expression of the two fusion protein populations in yeast are known and can be used (*see e.g.*, U.S. Patent No. 5,1468,614; Bartel et al., 1993, *Cellular Interactions in Development*, Hartley, ed., Practical Approach Series xviii, IRL Press at Oxford University Press, New York, NY, pp. 153-179; Fields and Sternglanz, 1994, Trends In Genetics 10:286-292). Any yeast strain or derivative strains made therefrom, known in the art can be used including N105, N106, N1051, N1061, and YULH. Other exemplary strains that can be used in the assay of the invention also include:

Y190: MATa, *ura3-52*, *his3-200*, *lys2-801*, *ade2-101*, *trpl-901*, *leu2-3,112*, gal4α, gal80α, *cyh'2*, *LYS2::GALl*$_{UAS}$*-HIS3*$_{TATA}$*HIS3*, *URA3::GALl*$_{UAS}$*-GALl*$_{TATA}$*-lacZ*; Harper et al., 1993, Cell 75:805-816, available from Clontech, Palo Alto, CA,. Y190 contains *HIS3* and *lacZ* reporter genes driven by GAL4 binding sites.

CG-1945: MATa, *ura3-52*, *his3-200*, *lys2-801*, *ade2-101*, *trpl-901*, *leu2-3,112*, *gal4-542*, *gal80-538*, *cyh'2*, *LYS2::GALl*$_{UAS}$*-HIS3*$_{TATA}$*HIS3*, *URA3::GALl*$_{UAS17mers(x3)}$*-CYC1*$_{TATA}$*-lacZ*, available from Clontech, Palo Alto, CA. CG-1945 contains *HIS3* and *lacZ* reporter genes driven by GAL4 binding sites.

Y187: MAT-α, *ura3-52*, *his3-200*, *ade2-101*, *trp1-901*, *leu2-3,112*, *gal4α*, *gal80α*, *URA3::GAL1*$_{UAS}$*-GAL1*$_{TATA}$*-lacZ*, available from Clontech, Palo Alto, CA. Y187 contains a *lacZ* reporter gene driven by GAL4 binding sites.

SFY526: MATa, *ura3-52*, *his3-200*, *lys2-801*, *ade2-101*, *trp1-901*, *leu2-3,112*, *gal4-542*, *gal80-538*, *can*^{r}, *URA3::GAL1-lacZ*, available from Clontech, Palo Alto, CA. SFY526 contains *HIS3* and *lacZ* reporter genes driven by GAL4 binding sites.

HF7c: MATa, *ura3-52*, *his3-200*, *lys2-801*, *ade2-101*, *trp1-901*, *leu2-3,112*, *gal4-542*, *gal80-538*, *LYS2::GAL1-HIS3*, *URA3::GAL1*$_{UAS\text{-}17MERS(x3)}$*-CYC1-lacZ*, available from Clontech, Palo Alto, CA. HF7c contains *HIS3* and *lacZ* reporter genes driven by GAL4 binding sites.

YRG-2: MATa, *ura3-52*, *his3-200*, *lys2-801*, *ade2-101*, *trp1-901*, *leu2-3,112*, *gal4-542*, *gal80-538*, *LYS2::GAL1*$_{UAS}$*-GAL1*$_{TATA}$*-HIS3*, *URA3::GAL1*$_{UAS17mers(x3)}$*-CYC1-lacZ*, available from Stratagene, La Jolla, CA. YRG-2 contains *HIS3* and *lacZ* reporter genes driven by GAL4 binding sites.

If not already lacking in endogenous reporter gene activity, cells mutant in the reporter gene may be selected by known methods, or the cells can be made mutant in the target reporter gene by known gene-disruption methods prior to introducing the reporter gene (Rothstein, 1983, Meth. Enzymol. 101:202-211).

In a specific embodiment, plasmids encoding the different fusion protein populations can be introduced simultaneously into a single host cell (*e.g.*, a haploid yeast cell) containing one or more reporter genes, by co-transformation, to conduct the assay for protein-protein interactions. Or, preferably, the two fusion protein populations are introduced into a single cell either by mating (*e.g.*, for yeast cells) or cell fusions (*e.g.*, of mammalian cells). In a mating type assay, conjugation of haploid yeast cells of opposite mating type that have been transformed with a binding domain fusion expression construct (preferably a plasmid) and an activation (or inhibitor) domain fusion expression construct (preferably a plasmid), respectively, will deliver both constructs into the same diploid cell. The mating type of a yeast strain may be manipulated by transformation with the HO gene (Herskowitz and Jensen, 1991, Meth. Enzymol. 194:132-146).

In a preferred embodiment, a yeast interaction mating assay is employed using two different types of host cells, strain-type a and alpha of the yeast *Saccharomyces cerevisiae*. The host cell preferably contains at least two reporter genes, each with one or more binding sites for the DNA-binding domain (*e.g.*, of a transcriptional activator). The activator domain and DNA binding domain are each parts of chimeric proteins formed from the two respective populations of proteins. One strain of host cells, for example the a strain, contains fusions of the library of nucleotide sequences with the DNA-binding domain of a transcriptional activator, such as GAL4. The hybrid proteins expressed in this set of host cells are capable of recognizing the DNA-binding site in the promoter or enhancer region in the reporter gene construct. The second set of yeast host cells, for example, the alpha strain, contains nucleotide sequences encoding fusions of a library of DNA sequences fused to the activation domain of a transcriptional activator.

In a preferred embodiment, the fusion protein constructs are introduced into the host cell as a set of plasmids. These plasmids are preferably capable of autonomous replication in a host yeast cell and preferably can also be propagated in *E. coli*. The plasmid contains a promoter directing the transcription of the DNA binding or activation domain fusion genes, and a transcriptional termination signal. The plasmid also preferably contains a selectable marker gene, permitting selection of cells containing the plasmid. The plasmid can be single-copy or multi-copy. Single-copy yeast plasmids that have the yeast centromere may also be used to express the activation and DNA binding domain fusions (Elledge et al., 1988, Gene 70:303-312).

The fusion constructs can be introduced directly into the yeast chromosome via homologous recombination mediated through yeast sequences that are not essential for vegetative growth of yeast, *e.g.*, the *MER2*, *MER1*, *ZIP1*, *REC102*, or *ME14* gene.

Bacteriophage vectors can also be used to express the DNA binding domain and/or activation domain fusion proteins. Libraries can generally be prepared faster and more easily from bacteriophage vectors than from plasmid vectors.

Methods can be used for detecting one or more protein-protein interactions comprising (a) recombinantly expressing a *C. elegans* insulin-like protein or a derivative or analog thereof in a first population of yeast cells being of a first mating type and comprising a first fusion protein containing the *C. elegans* insulin-like sequence and a DNA binding domain, wherein said first population of yeast cells contains a first nucleotide sequence operably linked to a promoter driven by one or more DNA binding sites recognized by said DNA binding domain such that an interaction of said first fusion protein with a second fusion protein, said second fusion protein comprising a transcriptional activation domain, results in increased transcription of said first nucleotide sequence; (b) negatively selecting to eliminate those yeast cells in said first population in which said increased transcription of said first nucleotide sequence occurs in the absence of said second fusion protein; (c) recombinantly expressing in a second population of yeast cells of a second mating type different from said first mating type, a plurality of said second fusion proteins, each second fusion protein comprising a sequence of a fragment, derivative or analog of a protein and an activation domain of a transcriptional activator, in which the activation domain is the same in each said second fusion protein; (d) mating said first population of yeast cells with said second population of yeast cells to form a third population of diploid yeast cells, wherein said third population of diploid yeast cells contains a second nucleotide sequence operably linked to a promoter driven by a DNA binding site recognized by said DNA binding domain such that an interaction of a first fusion protein with a second fusion protein results in

increased transcription of said second nucleotide sequence, in which the first and second nucleotide sequences can be the same or different; and (e) detecting said increased transcription of said first and/or second nucleotide sequence, thereby detecting an interaction between a first fusion protein and a second fusion protein.

In a preferred embodiment, the bait *C. elegans* insulin-like sequence and the prey library of chimeric genes are combined by mating the two yeast strains on solid media for a period of approximately 6-8 hours. Alternatively, the mating can be performed in liquid media. The resulting diploids contain both kinds of chimeric genes, *i.e.*, the DNA-binding domain fusion and the activation domain fusion.

Preferred reporter genes include the *URA3*, *HIS3* and/or the *lacZ* genes (*see e.g.*, Rose and Botstein, 1983, Meth. Enzymol. 101:167-180) operably linked to GAL4 DNA-binding domain recognition elements. Other reporter genes comprise the functional coding sequences for, but not limited to, Green Fluorescent Protein (GFP) (Cubitt et al., 1995, Trends Biochem. Sci. 20:448-455), luciferase, *LEU2*, *LYS2*, *ADE2*, *TRP1*, *CAN1*, *CYH2*, *GUS*, *CUP1* or chloramphenicol acetyl transferase (CAT). Expression of *LEU2*, *LYS2*, *ADE2* and *TRP1* are detected by growth in a specific defined media; *GUS* and *CAT* can be monitored by well known enzyme assays; and *CAN1* and *CYH2* are detected by selection in the presence of canavanine and cycloheximide. With respect to GFP, the natural fluorescence of the protein is detected, or a modified GFP having modified fluorescence is detected.

Transcription of the reporter gene can be detected by a linked replication assay. For example, as described by Vasavada et al., 1991, Proc. Natl. Acad. Sci. U.S.A. 88:10686-10690, expression of SV40 large T antigen is under the control of the E1B promoter responsive to GAL4 binding sites. The replication of a plasmid containing the SV40 origin of replication, indicates the reconstruction of the GAL4 protein and a protein-protein interaction. Alternatively, a polyoma virus replicon can be employed (Vasavada et al., 1991, Proc. Natl. Acad. Sci. U.S.A. 88:10686-10690).

The expression of reporter genes that encode proteins can also be detected using immunoassay methods. Alam and Cook (1990, Anal. Biochem. 188:245-254) disclose examples of detectable marker genes that can be operably linked to a transcriptional regulatory region responsive to a reconstituted transcriptional activator, and thus used as reporter genes.

The activation of reporter genes like *URA3* or *HIS3* enables the cells to grow in the absence of uracil or histidine, respectively, and hence serves as a selectable marker. Thus, after mating, the cells exhibiting protein-protein interactions are selected by the ability to

grow in media lacking a nutritional component, such as uracil or histidine (referred to as -URA (minus URA) and -HIS (minus HIS) medium, respectively). The -HIS medium preferably contains 3-amino-1,2,4-triazole (3-AT), which is a competitive inhibitor of the *HIS3* gene product, and thus, requires higher levels of transcription in the selection (*see* Durfee et al., 1993, Genes Dev. 7:555-569). Similarly, 6-azauracil, which is an inhibitor of the *URA3* gene product, can be included in -URA medium (Le Douarin et al., 1995, Nucl. Acids Res. 23:876-878). *URA3* gene activity can also be detected and/or measured by determining the activity of its gene product, orotidine-5'-monophosphate decarboxylase (Pierrat et al., 1992, Gene 119:237-245; Wolcott et al., 1966, Biochem. Biophys. Acta 122:532-534). In other embodiments of the present invention, the activities of the reporter genes like *GFP* or *lacZ* are monitored by measuring a detectable signal (*e.g.*, fluorescent or chromogenic, respectively) that results from the activation of these reporter genes. For example, *lacZ* transcription can be monitored by incubation in the presence of a chromogenic substrate, such as X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside), of its encoded enzyme, β-galactosidase. The pool of all interacting proteins isolated by this manner from mating the *C. elegans* insulin-like sequence product and the library identifies the "insulin-like interactive population".

False positives arising from transcriptional activation by the DNA binding domain fusion proteins in the absence of a transcriptional activator domain fusion protein can be prevented or reduced by negative selection for such activation within a host cell containing the DNA binding fusion population, prior to exposure to the activation domain fusion population. For example, if such cell contains *URA3* as a reporter gene, negative selection is carried out by incubating the cell in the presence of 5-fluoroorotic acid (5-FOA), which kills. Hence, if the DNA-binding domain fusions by themselves activate transcription, the metabolism of 5-FOA will lead to cell death and the removal of self-activating DNA-binding domain hybrids.

Negative selection involving the use of a selectable marker as a reporter gene and the presence in the cell medium of an agent toxic or growth inhibitory to the host cells in the absence of reporter gene transcription is preferred, since it allows a higher rate of processing than other methods. Negative selection can also be carried out on the activation domain fusion population prior to interaction with the DNA binding domain fusion population, by similar methods, either alone or in addition to negative selection of the DNA binding fusion population.

Negative selection can also be carried out on the recovered protein-protein complex by known methods (*see e.g.*, Bartel et al., 1993, BioTechniques 14:920-924) although pre-

negative selection (prior to the interaction assay) is preferred. For example, each plasmid encoding a protein (peptide or polypeptide) fused to the activation domain (one-half of a detected interacting complex) can be transformed back into the original screening strain, either alone or with a plasmid encoding only the DNA-binding domain, the DNA-binding domain fused to the detected interacting protein, or the DNA-binding domain fused to a protein that does not affect transcription or participate in the protein-protein interaction. A positive interaction detected with any plasmid other than that encoding the DNA-binding domain fusion to the detected interacting protein is deemed a false positive and is eliminated from the screen.

In a preferred embodiment, the *C. elegans* insulin-like plasmid population is transformed in a yeast strain of a first mating type (a or alpha), and the second plasmid population (containing the library of DNA sequences) is transformed in a yeast strain of a different mating type. Both strains are preferably mutant for *URA3* and *HIS3*, and contain *HIS3*, and optionally *lacZ*, as reporter genes. The first set of yeast cells are positively selected for the insulin-like plasmids and are negatively selected for false positives by incubation in medium lacking the selectable marker (*e.g.*, tryptophan) and containing 5-FOA. Yeast cells of the second mating type are transformed with the second plasmid population, and are positively selected for the presence of the plasmids containing the library of fusion proteins. Selected cells are pooled. Both groups of pooled cells are mixed together and mating is allowed to occur on a solid phase. The resulting diploid cells are then transferred to selective media that selects for the presence of each plasmid and for activation of reporter genes.

After an interactive population is obtained, the DNA sequences encoding the pairs of interactive proteins can be isolated by a method wherein either the DNA-binding domain hybrids or the activation domain hybrids are amplified, in separate respective reactions. Preferably, the amplification is carried out by polymerase chain reaction (PCR) using pairs of oligonucleotide primers specific for either the DNA-binding domain hybrids or the activation domain hybrids. This PCR reaction can also be performed on pooled cells expressing interacting protein complexes, preferably pooled arrays of interactants. Other amplification methods known in the art can be used, such as ligase chain reaction, use of Qβ replicase, or methods listed in Kricka et al., 1995, *Molecular Probing, Blotting, and Sequencing*, Academic Press, New York, Chapter 1 and Table IX.

The plasmids encoding the DNA-binding domain hybrid and the activation domain hybrid proteins can also be isolated and cloned by any known method. For example, if a shuttle (yeast to *E. coli*) vector is used to express the fusion proteins, the genes can be

recovered by transforming the yeast DNA into *E. coli* and recovering the plasmids from *E. coli*. Alternatively, the yeast vector can be isolated, and the insert encoding the fusion protein subcloned into a bacterial expression vector, for growth of the plasmid in *E. coli*.

Assays of insulin-like proteins

The functional activity of insulin-like proteins, derivatives and analogs can be assayed using known methods. For example, immunoassays can be used to test the ability to bind to an anti-insulin-like protein antibody, or to compete for binding with a wild-type insulin-like protein. Various competitive and non-competitive assay systems can be used such as radioimmunoassays, ELISA, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels), western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, *etc.* Physiological correlates of insulin-like protein binding to its substrates and/or receptors (*e.g.*, signal transduction) can be assayed.

In insect (*e.g.*, *D. melanogaster*), worm (*e.g.*, *C. elegans*), or other model systems, genetic studies can be done to study the phenotypic effect of an insulin-like gene mutant that is a derivative or analog of a wild-type insulin-like gene as described further below.

Antisense regulation of gene expression

The invention provides for antisense sequences of *C. elegans* insulin-like genes. An insulin-like "antisense" nucleic acid as used herein refers to a nucleic acid capable of hybridizing to a portion of an insulin-like RNA (preferably mRNA) by virtue of some sequence complementarity. Antisense nucleic acids may also be referred to as inverse complement nucleic acids. The antisense nucleic acid may be complementary to at least a portion of a coding and/or noncoding region of an insulin-like mRNA. Absolute complementarity is not required, but should be sufficient so that a stable duplex with the RNA can form. In the case of double-stranded insulin-like antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an insulin-like RNA it may contain and still form a stable duplex (or triplex, as the case may be). The degree of tolerable mismatch can be readily determined by calculating the melting point of the hybridized complex.

Antisense nucleic acids have utility in inhibiting an insulin-like protein function. For example, such antisense nucleic acids may be useful as pesticides to eradicate parasites in plants, or in animals such as dogs. A preferred antisense nucleic acid is a single stranded DNA oligonucleotide comprising a sequence antisense to the sequence encoding a B peptide domain or an A peptide domain of an insulin-like protein.

Preferably the antisense nucleic acids are oligonucleotides having at least 6 nucleotides and more preferably at least 10, 15, 20, or 50 nucleotides. Oligonucleotides having at least 100 or 200 nucleotides can also be used. The oligonucleotides can be double or single stranded RNA or DNA or chimeric mixtures or derivatives or modified versions thereof. One or more modifications can be made at the base or sugar moiety, or phosphate backbone. Examples of modified base moieties include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Examples of modified sugar moieties include arabinose, 2-fluoroarabinose, xylulose, and hexose. Examples of modifications at the phosphate backbone a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

The oligonucleotide may include other appending groups such as peptides, agents that facilitate transport across the cell membrane or blood-brain barrier, hybridization-triggered cleavage agents or intercalating agents.

The oligonucleotide can also be α-anomeric so that it forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other.

The oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization-triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, *etc.*

An insulin-like antisense oligonucleotide may comprises catalytic RNA, or a ribozyme (*see e.g.* WO 90/11364; Sarver et al., 1990, Science 247:1222-1225). In another embodiment, the oligonucleotide is a 2′-0-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

The oligonucleotides may be synthesized by known methods, *e.g.*, by use of an automated DNA synthesizer (commercially available from Biosearch, Applied Biosystems, *etc.*). Phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16:3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), *etc.* Alternatively, the insulin-like antisense nucleic acids can be produced intracellularly by transcription from an exogenous sequence. For example, a vector can be introduced *in vivo* such that it is taken up by a cell, within which cell the vector or a portion thereof is transcribed, producing an antisense nucleic acid (RNA). Such a vector would contain a sequence encoding the insulin-like antisense nucleic acid. The vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the insulin-like antisense RNA can be by any promoter, inducible or constitutive, known to act in mammalian cells, such as those previously discussed.

Identifying signaling pathways and phenotypes

Animal models which may be used in the identification and characterization of *C. elegans* insulin-like protein signaling pathways, and/or phenotypes associated with the mutation or abnormal expression of a *C. elegans* insulin-like protein. Methods of producing a variety of animal models using novel genes and proteins are well known (*see e.g.*, WO 96/34099); three examples are discussed below.

In one type of animal model a normal *C. elegans* insulin-like gene has been recombinantly introduced into the genome of the animal as an additional gene, under the regulation of either an exogenous or an endogenous promoter element, and as either a minigene or a large genomic fragment. The normal gene can be recombinantly substituted (e.g. by homologous recombination or gene targeting) for one or both copies of the animal's homologous gene.

In a second model animal, a mutant *C. elegans* insulin-like gene has been recombinantly introduced into the genome of the animal as an additional gene, under the regulation of either an exogenous or an endogenous promoter element, and as either a minigene or a large genomic fragment. The mutant gene can be recombinantly substituted for one or both copies of the animal's homologous gene.

Third, animals are provided in which a mutant version of one of that animal's own genes (bearing, for example, a specific mutation corresponding to, or similar to, a pathogenic mutation of an insulin-like gene from another species) has been recombinantly introduced into the genome of the animal as an additional gene, under the regulation of either an exogenous or an endogenous promoter element, and as either a minigene or a large genomic fragment.

Finally, equivalents of transgenic animals, including animals with mutated or inactivated genes, may be produced using chemical or x-ray mutagenesis. Using the isolated nucleic acids disclosed herein one may more rapidly screen the resulting offspring by, for example, direct sequencing, restriction fragment length polymorphism (RFLP) analysis, PCR, or hybridization analysis to detect mutants, or Southern blotting to demonstrate loss of one allele.

Such animal models may be used to identify phenotypes associated with mutation or abnormal expression of a *C. elegans* insulin-like protein and to identify a *C. elegans* insulin-like protein signaling pathway. For example, a *C. elegans* insulin-like gene can be disrupted (e.g mutated or abnormally expressed) and the effect can be identified using any suitable assay commonly used in *C. elegans* research (e.g. a dauer formation assay, a developmental assay, an energy metabolism assay, a growth rate assay and a reproductive capacity assay). The gene can be disrupted by any suitable method such as EMS chemical deletion mutagenesis, transposon insertion mutagenesis, or double-stranded RNA interference, as discussed in detail below.

Abnormal expression can be overexpression, underexpression (*e.g.*, due to inactivation), expression at a developmental time different from wild-type animals, or expression in a cell type different from in wild-type animals.

Assays for changes in gene expression

Changes in the expression of identified *C. elegans* insulin-like genes and proteins can be detected using known (*see e.g.*, WO 96/34099). Such assays may be performed *in vitro* using transformed cell lines, immortalized cell lines, or recombinant cell lines, or *in vivo* using animal models. The assays may detect the presence of increased or decreased

expression of a *C. elegans* insulin-like gene or protein on the basis of increased or decreased mRNA expression (using, *e.g.*, nucleic acid probes), increased or decreased levels of related protein products (using, *e.g.*, the antibodies disclosed herein), or increased or decreased levels of expression of a marker gene (*e.g.*, β-galactosidase or luciferase) operably linked to a 5' regulatory region in a recombinant construct.

Various expression analysis techniques may be used to identify genes which are differentially expressed between two conditions, such as a cell line or animal expressing a normal *C. elegans* insulin-like gene compared to another cell line or animal expressing a mutant *C. elegans* insulin-like gene. Such techniques include differential display, serial analysis of gene expression (SAGE), nucleic acid array technology, subtractive hybridization, proteome analysis and mass-spectrometry of two-dimensional protein gels. Nucleic acid array technology (*i.e.*, gene chips) may be used to determine a global (*i.e.*, genome-wide) gene expression pattern in a normal *C. elegans* animal for comparison with an animal having a mutation in one or more *C. elegans* insulin-like genes.

Gene expression profiling can be used to identify other genes (or proteins) that may have a functional relation to (*e.g.*, may participate in a signaling pathway with) a *C. elegans* insulin-like gene. The genes are identified by detecting changes in their expression levels following mutation, *i.e.*, insertion, deletion or substitution in, or overexpression, underexpression, mis-expression or knock-out, of a *C. elegans* insulin-like gene, as described in the examples below. Expression profiling methods provide a powerful approach for analyzing the effects of mutation in a *C. elegans* insulin-like gene. A variety of methods are well known in the art including subtractive hybridization, differential display, serial analysis of gene expression (SAGE), proteome analysis, and hybridization-based methods employing nucleic acid arrays.

Identification of compounds with binding capacity

Screening methodologies can be used for the identification of proteins and other compounds which bind to, or otherwise directly interact with, the *C. elegans* insulin-like genes and proteins of the invention. Suitable screening methods are disclosed in WO 96/34099. The proteins and compounds include endogenous cellular components which interact with the identified genes and proteins *in vivo* and which, therefore, may provide new targets for pharmaceutical and therapeutic interventions, as well as recombinant, synthetic, and otherwise exogenous compounds which may have binding capacity and, therefore, may be candidates for pharmaceutical agents. Thus, cell lysates or tissue homogenates may be screened for proteins or other compounds which bind to one of the

normal or mutant *C. elegans* insulin-like genes and proteins. Alternatively, any of a variety of exogenous compounds, both naturally occurring and/or synthetic (*e.g.*, libraries of small molecules or peptides), may be screened for binding capacity. Typically, a screening method comprises the step of mixing a *C. elegans* insulin-like protein or fragment or derivative thereof with test compounds, allowing time for any binding to occur, and assaying for any bound complexes.

EXAMPLES

The following examples are provided merely as illustrative of various aspects of the invention and shall not be construed to limit the invention in any way. The Examples describe the discovery of an unexpectedly large family of insulin-like genes in *C. elegans* which includes the 31 genes as illustrated in the alignment of FIG. 3 and in FIGs 4-36 and described in detail below. The SEQ ID NO for each protein and cDNA corresponding to these insulin-like genes is set forth in Table 1 below.

Table 1. *C. elegans* insulin-like genes and the corresponding sequence identification number (SEQ ID NO:) for each encoded protein and cDNA. See FIG. 4 through FIG. 34 for annotated sequences.

	SEQ ID NO:	
gene	protein	cDNA
F13B12.N	1	19
ZK75.1	2	20
ZK75.2	3	21
ZK75.3	4	22
ZK84.6	5	23
ZK84.N2	6	24
ZK1251.2	7	25
ZK1251.N	8	26
C06E2.N	9	27
C17C3.4	10	28
C17C3.N	11	29
M04D8.1	12	30
M04D8.2	13	31
M04D8.3	14	32
ZK84.N	15	33

F56F3.6	16	34
T28B8.N	17	35
ZC334.N	18	36
T08G5.N	158	162
F41G3.N	159	163
F41G3.N2	160	164
C17C3.N2	161	165
ZC334.N2	198	207
ZC334.N3	199	208
ZC334.N4	200	209
ZC334.N5	201	210
ZC334.N6	202	211
ZC334.N7	203	212
T10D4.N	204	213
T10D4.N2	205	214
Y52A1.N	206	215

EXAMPLE 1: PCR CLONING OF *C. ELEGANS* INSULIN-LIKE cDNAs

Twenty-two *C. elegans* insulin-like genes have been cloned using the polymerase chain reaction (PCR), as described in detail below. See Table 1 for the assigned name of each of the eighteen *C. elegans* insulin-like genes, and the corresponding sequence identification number for the nucleotide sequence of each cDNA and the amino acid sequence of each protein.

PCR primers were designed for cloning each gene under the following general rationale. For further details specific for each gene, see the Examples section below.

Genes ZK75.3, ZK75.1, ZK1251.2 and ZK1251.N were all predicted to have an SL1 splice acceptor upstream of the predicted start codon. Therefore, the SL1 sequence was used as the upstream primer for each of these cDNAs. ZK84.6 was predicted to have a splice acceptor upstream of the start codon; however, no PCR product was obtained using SL1 as an upstream primer. Therefore, the sequence immediately following the predicted splice acceptor was used. The downstream primers were chosen to fall downstream of the predicted stop codon.

For M04D8.1, M04D8.2, M04D8.3, C17C3.4, C17C3.N, F13B12.N, T28B8.N, ZC334.N, and ZK84.N, primers had a HindIII site on the end of the 5' primer according to

the formula CCC-AAGCTT-N, where N = 24 to 26 specific nucleotides; and an XbaI site on the end of the 3' primer according to the formula GC-TCTAGA-N, where N = 24 to 26 specific nucleotides. The engineered restriction sites of these primers were used for cloning. F56F3.6 has an internal XbaI site, so an XhoI site was used instead on the 3' primer. What follows is a list of conditions used for PCR amplification and cloning of each gene.

ZK75.1

The template DNA source was a mixed-stage, *C. elegans* cDNA library, oligo-dT primed and ligated into UniZap XR (phage lambda) vector available from Stratagene. The library DNA was prepared by Qiagen purification and adjusted to a concentration of 70 ng/µl.

The cDNA was generated by the polymerase chain reaction (PCR) procedure, using the Boehringer Mannheim Expand High Fidelity PCR System. Each reaction was performed in a total volume of 100 µl. The components of the reaction were 1 µl (70 ng) template DNA, 200 µM each dNTP, 300 nM each primer as described below, 1X buffer with $MgCl_2$ as supplied by the manufacturer, and 2.6 U of enzyme.

First, the primers were pooled and denatured at 95°C for 5:00 (where 0:00 indicates time in minutes:seconds), and stored on ice. The remainder of the reaction mixture was added, and the PCR reaction started as follows:

95°C for 2:00
35 cycles of: 95°C for 0:15
54°C for 0:30
72°C for 1:00
72° for 5:00

For the first round of PCR, the primers used were as follows:

75.1 GACGGAGATGGCTTGTTGGACGAC (SEQ ID NO:37)
SL1 GGTTTAATTACCCAAGTTTGAG (SEQ ID NO:38)

The first round of PCR yielded no detectable band as determined by agarose gel electrophoresis, staining with ethidium bromide, and visualization on a long-wave UV light box. Accordingly, a second round of PCR was next performed as described above, except with the following changes. The template DNA was 1 µl of the first round PCR reaction, the reactions were run for 20 cycles only, and different (nested) primers were used as follows:

75.1.5' CAAGAGAATGTTTTCATTCTTTAC (SEQ ID NO:39)
75.1B TTACTTTTCTGGGCAGCAAGCTTG (SEQ ID NO:40)

The second PCR reaction yielded a strong single band of DNA at the predicted size. To subclone this PCR product into a plasmid vector for DNA sequencing, we first isolated the PCR product by agarose gel electrophoresis (90 µl of the second PCR reaction run on a 1.2% gel). We excised the band with a razor blade and purified the product from the gel using the Prep-a-Gene kit from BioRad. We then ligated the PCR product into the plasmid vector PCRII and transformed *E. coli* using an InVitrogen TA Cloning Kit. We screened bacterial colonies for the correct plasmid by preparing mini-prep DNA using the Primm Labs Mini-Prep kit, and analyzed the mini-prep DNA by EcoRI restriction digest and agarose gel electrophoresis.

We sequenced the subcloned PCR products by thermal cycling, using the Big Dye ready reaction mix sequencing kit. For each sequencing reaction, we added: approximately 100 ng of mini-prep DNA; 0.8 pmol of sequencing primer; 1.5 µl 5X Big Dye ready reaction buffer; 1 µl 80 mM Tris, 2 mM $MgCl_2$ pH 9.0; and adjusted the volume to 10 µl with distilled water. The M13 Forward and M13 Reverse sequencing primers were used. The sequencing reactions were thermal cycled using the following program:

96° for 5:00
25 cycles of: 96°C for 0:30
50°C for 0:1560°C for 4:00

We precipitated the cycled DNA with 75 µl 70% ethanol/5 mM $MgCl_2$ by incubating at room temperature for 20 minutes. We recovered the precipitated DNA by centrifugation at 15,000 X g for 30 minutes, removed the supernatant, and further dried the DNA pellet by vacuum centrifugation for 10 minutes. The sequencing reactions were analyzed and the DNA sequence determined by gel electrophoresis and fluorescent detection of sequencing products.

ZK75.2

The template DNA source was mixed-stage *C. elegans* first strand cDNA, poly-A selected and oligo-dT primed using the Gibco-BRL Superscript kit. The RNA was removed by RNAse digestion, and the cDNA was diluted with TE buffer and adjusted to a final concentration of approximately 70 ng/µl. The cDNA was generated by the polymerase chain reaction (PCR) procedure, using the Boehringer Mannheim Expand High Fidelity PCR System. Each reaction was performed in a total volume of 100 µl. The components of the reaction were 1 µl (70 ng) template DNA, 200 µM each dNTP, 300 nM each primer as described below, 1X buffer with $MgCl_2$ as supplied by the manufacturer, and 2.6 U enzyme.

First, the template was denatured at 95°C for 5:00 minutes and stored on ice. The remainder of the reaction mixture was added, and the PCR reaction started as follows:

95°C for 2:00

35 cycles of: 95°C for 0:15

54°C for 0:30

72°C for 1:00

72° for 5:00

For the first round of PCR, the primers were as follows:

75.2.5' CTACCATGAACGCTATAATCTTCT (SEQ ID NO:41)

75.2.3' ATGATAGTACGATATGTCCATAAC (SEQ ID NO:42)

This reaction yielded a single strong band of the expected size (349 bp) after one round of PCR.

To subclone the PCR product into a plasmid vector for DNA sequencing, we first isolated the PCR product by agarose gel electrophoresis (90 µl of the second PCR reaction run on a 1.2% gel). We excised the band with a razor blade, and purified the product from the gel using the Prep-a-Gene kit from BioRad. We then ligated the PCR product into the plasmid vector PCRII and transformed *E. coli* using the InVitrogen TA Cloning Kit. We screened bacterial colonies for the correct plasmid by colony PCR, using the following primers:

75.2.5' CTACCATGAACGCTATAATCTTCT (SEQ ID NO:41)

75.2.3' ATGATAGTACGATATGTCCATAAC (SEQ ID NO:42)

To confirm the positive colonies, we prepared mini-prep plasmid DNA from positive colonies using the Primm Labs miniprep kit and confirmed the plasmid by EcoR1 restriction digest and agarose gel electrophoresis.

We analyzed the sequence of the PCR product as described for ZK75.1.

ZK75.3

A first round PCR reaction was performed exactly as for ZK75.2, except using primers:

75.3 CCTATTTTCCAGCCACAGCACTCTC (SEQ ID NO:43)

SL1 GGTTTAATTACCCAAGTTTGAG (SEQ ID NO:38)

No band was obtained after the first round of PCR. Strong bands of 426 bp were obtained after the second round of PCR, which was performed as follows:

template = 2 µl of first round PCR

same primers as first round

same PCR conditions as first round

Subcloning and sequencing of the second round reaction product was performed exactly as for ZK75.1.

ZK84.6

First round PCR was performed exactly as for ZK75.1, except using primers:

84.3OUTER CCCCGTACTCATTTTCCGTTATCC (SEQ ID NO:44)

84.3 GTATGGTACAGAGACTGATATCGG (SEQ ID NO:45)

A strong single band of 423 bp after the first round of PCR was obtained.

Subcloning and sequencing of PCR products was performed exactly as for ZK75.2, except using the following primers for colony PCR screening:

84.3OUTER CCCCGTACTCATTTTCCGTTATCC (SEQ ID NO:44)

84.3.5'B CAAGGAAAATGCACTCGATCGTCG (SEQ ID NO:46)

ZK84.N

The template DNA source was a mixed stage *C. elegans* cDNA library oligo primed and ligated into UniZap XR (phage lambda) vector, purchased from Stratagene. The library DNA was prepared by Qiagen purification and adjusted to a concentration of 70 ng/µl.

The cDNA was generated by the polymerase chain reaction (PCR) procedure, using the Boehringer Mannheim Expand High Fidelity PCR System. Each reaction was performed in a total volume of 50 µl. The components of the reaction were 0.5 µl (70 ng) template DNA, 100 µM each dNTP, 150 nM each primer as described below, 1X buffer with $MgCl_2$ as supplied by the manufacturer, and 1.3 U enzyme.

First, the template was denatured at 95°C for 5:00 minutes, and stored on ice. The remainder of the reaction mixture was added, and the PCR reaction started as follows:

95°C for 2:00

35 cycles of: 95°C for 0:15

54°C for 0:30

72°C for 1:00

For the first round of PCR, the primers were:

84.NF-HIN CCCAAGCTTTGTTATTTAATGATGTGGAGATGG (SEQ ID NO:47)

84.NR-XBA GCTCTAGAATGGTAAATACAGAACATTGGTTC (SEQ ID NO:48)

This reaction yielded a strong single band of DNA at the predicted size. To subclone the PCR product into a plasmid vector for DNA sequencing, we first purified the PCR product with the Geneclean kit (Bio101), then digested the product with HindIII and

XbaI and isolated the PCR product by agarose gel electrophoresis (45 µl of the PCR reaction run on a 1.2% gel). We excised the band with a razor blade, and purified the product from the gel using the Geneclean kit. We then ligated the cut PCR product into the plasmid vector pcDNA3.1 (InVitrogen) cut with HindIII/XbaI and transformed *E. coli*. We screened bacterial colonies for the correct plasmid by preparing mini-prep DNA using the Primm Labs Mini-Prep kit, and analyzed the mini-prep DNA by PmeI restriction digest and agarose gel electrophoresis.

We sequenced the subcloned PCR products by thermal cycling, using the Big Dye ready reaction mix sequencing kit. For each sequencing reaction, we added approximately 100 ng of mini-prep DNA; 0.8 pmol of sequencing primer; 1 µl 5X Big Dye ready reaction buffer; 1.5 µl 80 mM Tris, 2 mM $MgCl_2$, pH 9.0; and adjusted the volume to 10 µl with distilled water. The sequencing primers used were pcDNA3.1BGHReverse and a T7 promoter primer. The sequencing reactions were thermal cycled using the following program:

96° for 5:00
25 cycles of: 96°C for 0:30
50°C for 0:15
60°C for 4:00

We precipitated the cycled DNA with 75 µl 70% ethanol/5 mM $MgCl_2$ by incubating at room temperature for 20 minutes. We recovered the precipitated DNA by centrifugation at 15,000 X g for 30 minutes, removed the supernatant, and further dried the DNA pellet by vacuum centrifugation for 10 minutes. The sequencing reactions were analyzed and the DNA sequence determined by gel electrophoresis and fluorescent detection of sequencing products.

ZK84.N2

PCR was performed exactly as for ZK84.N, except using PCR primers:
ORPR-XBA GCTCTAGAGTGACGGTAGGTGTGTAGATGAAC (SEQ ID NO:49)
84.35' ATCGAAACTCTTCAATCTTCAAGG (SEQ ID NO:50)

This reaction yielded a strong single band of DNA at the predicted size. To subclone the PCR product into a plasmid vector for DNA sequencing, we first isolated the PCR product by agarose gel electrophoresis (45 µl of the PCR reaction run on a 1.2% gel). We excised the band with a razor blade, and purified the product from the gel using the Geneclean kit. We then ligated the PCR product into the plasmid vector PCRII and transformed *E. coli* using the InVitrogen TA Cloning Kit. We screened bacterial colonies

for the correct plasmid by preparing mini-prep DNA using the Primm Labs MiniPrep kit, and analyzed the mini-prep DNA by PmeI restriction digest and agarose gel electrophoresis.

We sequenced the subcloned PCR products by thermal cycling, using the Big Dye ready reaction mix sequencing kit. For each sequencing reaction, we added approximately 100 ng of mini-prep DNA; 0.8 pmol of sequencing primer; 1 µl 5X Big Dye ready reaction buffer; 1.5 µl 80mM Tris, 2 mM $MgCl_2$, pH 9.0; and adjusted the volume to 10 µl with distilled water. The sequencing primers used were pcDNA3.1BGHReverse and a T7 promoter primer. The sequencing reactions were thermal cycled using the following program:

96° for 5:00
25 cycles of: 96°C for 0:30
50°C for 0:15
60°C for 4:00

We precipitated the cycled DNA with 75 µl 70% ethanol/5 mM $MgCl_2$ by incubating at room temperature for 20 minutes. We recovered the precipitated DNA by centrifugation at 15,000 x g for 30 minutes, removed the supernatant, and further dried the DNA pellet by vacuum centrifugation for 10 minutes. The sequencing reactions were analyzed and the DNA sequence determined by gel electrophoresis and fluorescent detection of sequencing products.

ZK1251.2

PCR was performed exactly as for ZK75.1, except using primers:

SL1 GGTTTAATTACCCAAGTTTGAG (SEQ ID NO:38)
1251.2 GATAGAAGAAATTAAGGACAGCAC (SEQ ID NO:51)

A single strong band of 351 bp was obtained after one round of PCR. Subcloning and sequencing of PCR products was performed exactly as for ZK75.1.

ZK1251.N

PCR was performed exactly as for ZK75.1, except using primers:

1251.N GTAAACGATTAGATTAAGGACAAC (SEQ ID NO:52)
SL1 GGTTTAATTACCCAAGTTTGAG (SEQ ID NO:38)

No band was obtained after the first round of PCR. A second round was performed using an aliquot of the first round reaction as template, the same reaction mix and primers, and the same PCR conditions. Strong bands of 349 bp were obtained after the second round of PCR. Subcloning and sequencing was performed exactly as for ZK75.1.

C06E2.N

PCR was performed exactly as for ZK75.1, except using primers:

C06E2.5' GAGGAGTGAAACGATGATCGTCAC (SEQ ID NO:53)

C06E2 ATCCAATTGAGAAGACGATTGTTG (SEQ ID NO:54)

No band was obtained after the first round of PCR. A second round of PCR was performed using an aliquot of the first round as template, the same reaction mix and primers, and the same PCR conditions as in the first round, but for 20 cycles rather than 35 cycles.

A single strong band of 404 bp was obtained after the second round of PCR. Subcloning and sequencing of PCR products was performed exactly as ZK75.1.

M04D8.1

PCR was performed exactly as for ZK84.N, except using primers:

8.1F-HIN CCCAAGCTTTTGAACCATGAAAACCTACTCATT (SEQ ID NO:55)

8.1R-XBA GCTCTAGAGCTTTTTTTTATTCGGGACAGCAA (SEQ ID NO:56)

M04D8.3

PCR was performed exactly as for ZK84.N, except using primers:

8.3F-HIN CCCAAGCTTGGATTTCTGGAATTTCGATAATG (SEQ ID NO:57)

8.3R-XBA GCTCTAGAGCAGCATAGAATGGCGGAAGATC (SEQ ID NO:58)

C17C3.4

PCR was performed exactly as for ZK84.N, except using primers:

3.4F-HIN CCCAAGCTTGTGTAGGAATCGTTAAATATGTCT (SEQ ID NO:59)

3.4R-XBA GCTCTAGAGAGATCATATTATATTACACGAAC (SEQ ID NO:60)

F13B12.N

PCR was performed exactly as for ZK84.N, except using primers:

B12F-HIN CCCAAGCTTCCGCTCTCAACAACGGGCCACACG (SEQ ID NO:61)

B12R-XBA GCTCTAGAGATGAATAAGTTATCAATTATCGT (SEQ ID NO:62)

T28B8.N

PCR was performed exactly as for ZK84.N, except using primers:

SL1-HIN CCCAAGCTTGGTTTAATTACCCAAGTTTGAG (SEQ ID NO:63)

B8.2R-XBA GCTCTAGATGATGCGTATTTTGTGGGCGGTAC (SEQ ID NO:64)

ZC334.N

PCR was performed exactly as for ZK84.N, except using primers:

SL1-HIN CCCAAGCTTGGTTTAATTACCCAAGTTTGAG (SEQ ID NO:63)

34.NR-XBA GCTCTAGACTCATCAGTTGAAAATGAATTTAAG (SEQ ID NO:65)

F36F3.6

PCR was performed exactly as for ZK84.N, except using primers:

F3.6F-HIN CCCAAGCTTGGCATAAGCGAGTATCTGTGATCC (SEQ ID NO:66)

F3.6R-XHO CCGCTCGAGGTAAAGCGAGGGTAAAGTAGATCG (SEQ ID NO:67)

M04D8.2

PCR was performed exactly as for ZK84.N, except using primers:

8.2F-HIN CCCAAGCTTCTAACCAACAAAAATGCACACTAC (SEQ ID NO:68)

8.2R-XBA GCTCTAGACACGTGAACAATCTTTATCTTTAT (SEQ ID NO:69)

C17C3.N

PCR was performed exactly as for ZK84.N, except using primers:

3.NF-HIN CCCAAGCTTCACAGCCAAAAACAAAAATGCAATC (SEQ ID NO:70)

3.NR-XBA GCTCTAGACACAGTATTTTAATGAAGGAGATC (SEQ ID NO:71)

T08G5.N

PCR was performed exactly as for ZK84.N, except using 0.5 µl (35 ng) of template DNA and PCR primers:

SL1-HIN CCCAAGCTTGGTTTAATTACCCAAGTTTGAG (SEQ ID NO:144)

G5.NR-XBA GCTCTAGATAATTCAATGAAAAGGCAAAACGACG (SEQ ID NO:145)

This reaction yielded four bands after one round of PCR. The cDNA was contained within an approximately 315 bp DNA fragment. Subcloning and sequencing of PCR products was performed exactly as for ZK75.1 except with the following sequencing primers:

pcDNA3.1BGH Reverse TAGAAGGCACAGTCGAGG (SEQ ID NO:146)
T7 promoter primer TAATACGACTACTATAGGG (SEQ ID NO:147)

F41G3.N

PCR was performed exactly as for TO8G5.N, except using PCR primers:
G3.NF-HIN CCCAAGCTTCTTCATTTGGGCTTCATTTTACCAC (SEQ ID NO:148)
G3.NR-XBA GCTCTAGAGAAACAATGTTTTTATTCAACATG (SEQ ID NO:149)

This reaction yielded a band of the expected size after one round of PCR. The PCR product was cloned into pcDNA3.1 and sequenced exactly as described for ZK75.1.

F41G3.N2

PCR was performed exactly as for TO8G5.N, except using PCR primers:
G3.N2F-OUT CCCAAGCTTGGACTTTATCACAATTTCCAGCAC (SEQ ID NO:154)
G3.N2R-XBA GCTCTAGAGTTTCTAGATTTTTAGATTTCGTG (SEQ ID NO:155)
No band was visualized after the first round of PCR. A second PCR was performed as described above with the following changes: the template DNA was 1 of the first round PCR reacton, the reactions were run for 20 cycles only, and a different (nested) 3' primer was used. The primers were:
G3.N2F-XHO CCGCTCGAGATAATGAAGCTTCTTCTTCTCATTG (SEQ ID NO:156)
G3.N2R-XBA GCTCTAGAGTTTCTAGATTTTTAGATTTCGTG (SEQ ID NO:157)

This reaction yielded a band of the expected size. The PCR product was subcloned into pcDNA3.1 and sequenced exactly as described for TO85G.N, except the restriction enzymes used to digest the PCR product and vector were XbaI and XhoI.

C17C3.N2

PCR was performed exactly as for TO8G5.N, except using PCR primers:
C3.N2F-XHO CCGCTCGAGCTCGACGTTCTTCAATCTATATTTC (SEQ ID NO:150)
C3.N2R-XBA GCTCTAGACAAACACCATTAAATCTGTATTTAAAC (SEQ ID NO:151)

No band appeared after the first round of PCR. A second round of PCR was performed exactly as before using the following primers:
C3N2F-XHO CCGCTCGAGCTCGACGTTCTTCAATCTATATTTC (SEQ ID NO:164)
C3.N2R-INN GCTCTAGAGTTCACAAATTCATTTTCAAATACG (SEQ ID NO:165)

This reaction yield a single strong band of the expected size. The PCR product was subcloned into pcDNA3.1 and sequenced exactly as described for TO8G5.N, except the restriction enzymes used to digest the PCR product and vector were XbaI and XhoI.

Y52A1.N

The template DNA source was mixed-stage *C. elegans* first-strand cDNA, poly-A selected and oligo-dT primed using the Gibco-BRL Superscript kit. The RNA was removed by RNAse digestion, and the cDNA was diluted with TE buffer and adjusted to a final concentration of approximately 70 ng/µl.

The cDNA was generated by the polymerase chain reaction (PCR) procedure, using the Boehringer Mannheim Expand High Fidelity PCR System. Each reaction was performed in a total volume of 50 µl. The components of the reaction were 0.5 µl (35ng) template DNA, 100 µM each dNTP, 150 nM each primer as described below, 1X buffer with $MgCl_2$ as supplied by the manufacturer, and 1.3 units of enzyme.

First, the template was denatured at 95°C for 5:00 minutes, and stored on ice. The remainder of the reaction mixture was added, and the PCR reaction started as follows:

95°C for 2:00
35 cycles of: 95°C for 0:15
54°C for 0:30
72°C for 1:00

For the first round of PCR, the primers were:

SL1-HIN CCCAAGCTTGGTTTAATTACCCAAGTTTGAG (SEQ ID NO:166)
A1.1R-XBA GCTCTAGACAATTTTGATATTAAATTTTGTCG (SEQ ID NO:167)

The first round of PCR yielded no detectable band as determined by agarose gel electrophoresis, staining with ethidium bromide, and visualization on a UV light box.

A second round of PCR was performed as described above, with the following changes: the template DNA was 1 µl of the 1st round PCR reaction, the reactions were run for 20 cycles only, and a different (nested) 3' primer was used. The primers were:

SL1-HIN CCCAAGCTTTGGTTTAATTACCCAAGTTTGAG (SEQ ID NO:168)
1.1R-INN GCTCTAGATAAATTTTGTCGATTTTCAAGTTG (SEQ ID NO:169)

This reaction yielded a strong single band of DNA at approximately 1.3 kb.

To subclone the PCR product into a plasmid vector for DNA sequencing, we first isolated the PCR product by agarose gel electrophoresis (45µl of the second PCR reaction run on a 1.2% gel). We excised the band with a razor blade, and purified the product from the gel using the Geneclean (Bio101). We then ligated the PCR product into the plasmid

vector pCRII and transformed *E. coli* using the InVitrogen TA Cloning Kit. We screened bacterial colonies for the correct plasmid by preparing mini-prep DNA (Biotechniques 8, 172-3), and analyzed the mini-prep DNA by EcoRI restriction digest and agarose gel electrophoresis.

We sequenced the subcloned PCR products by thermal cycling, using the Big Dye ready reaction mix sequencing kit. For each sequencing reaction, we added: approximately 100 ng of mini-prep DNA; 0.8 pmol of sequencing primer; 1 µl 5X BigDye ready reaction buffer; 1.5 µl 80 mM Tris, 2 mM $MgCl_2$, pH 9.0; and adjusted the volume to 10 µl with distilled water. The following sequencing primers were used:

M13 Forward GTTTTCCCAGTCACG (SEQ ID NO:170)

M13 Reverse CAGGAAACAGCTATGAC (SEQ ID NO:171)

The sequencing reactions were thermal cycled using the following program:

96° for 5:00

25 cycles of: 96°C for 0:30

50°C for 0:15

60°C for 4:00

We precipitated the cycled DNA with 75 µl 70% ethanol/ 5 mM $MgCl_2$ by incubating at room temperature for 20 minutes. We recovered the precipitated DNA by centrifugation at 15,000 X g for 30 minutes, removed the supernatant, and further dried the DNA pellet by vacuum centrifugation for 10 minutes. The sequencing reactions were analyzed and the DNA sequence determined by gel electrophoresis and fluorescent detection of sequencing products. The resulting DNA sequence for the Y52A1-derived product indicated that there were in fact two opening reading frames in this cDNA. The open reading frame closest to the 5'-end of the message corresponding to this cDNA was not related to the insulin family. Instead, the insulin-like sequences predicted from the search of genomic DNA were found to correspond to the second open reading frame of this mRNA. Comparison of this Y52A1-derived cDNA sequence with the genomic sequence suggested that the likely explanation for this configuration of two open reading frames was that they correspond to an operon where multiple mRNAs are derived from the same transcription unit through different patterns of trans-splicing (see Zorio et al., 1994, Operons as a common form of chromosomal organization in *C. elegans*, Nature 372, 270-272). Thus, it was assumed that the insulin-like open reading frame in the Y52A1-derived product is actually translated from an mRNA that may be generated using an alternative trans-spliced leader such as SL2 or other leaders related to SL2.

PCR was used to amplify the presumptive insulin-like coding region from the larger cDNA product derived above. PCR was performed as above, with the following changes: the template was 1 µl of mini-prep DNA, and the following program was used:

95°C for 2:00

10 cycles of: 95°C for 0:30

54°C for 0:30

72°C for 1:00

The primers were:

Y52A1-i CCCAAGCTTGAGCATTTTGTTGCTCTGCAAAATG (SEQ ID NO:172)

1.1R-INN GCTCTAGATTAAATTTTGTCGATTTCAAGTTG (SEQ ID NO:173)

This reaction yielded a 268 bp product.

To subclone the PCR product into a plasmid vector for DNA sequencing, we first purified the PCR product with the Geneclean kit (Bio101), then digested the product with HindIII and XbaI and isolated the PCR product by agarose gel electrophoresis (45 µl of the PCR reaction run on a 1.2% gel). We excised the band with a razor blade, and purified the product from the gel using the Geneclean kit. We then ligated the cut PCR product into the plasmid vector pcDNA3.1 (InVitrogen) cut with HindIII/XbaI and transformed *E. coli*. We screened bacterial colonies for the correct plasmid by preparing mini-prep DNA (Biotechniques 8, 172-3), and analyzed the mini-prep DNA by PmeI restriction digest and agarose gel electrophoresis.

We sequenced the subcloned PCR products exactly as above, except with the following sequencing primers:

pcDNA3.1BGH Reverse TAGAAGGCACAGTCGAGG (SEQ ID NO:174)

T7 promoter primer TAATACGACTACTATAGGG (SEQ ID NO:175)

ZC334.N2

The cloning sites, HindIII and XbaI were used for many of the cDNAs except ZC334.N2, which has internal HindIII and XbaI sites. The 5' primer contains a BamHI restriction site on the 5' end: CG-GGATCC-N=24; and the 3' primer contains an EcoRI site on the end: CG-GAATTC-N=25.

The template DNA source was mixed stage *C. elegans* first strand cDNA, poly-A selected and oligo-dT primed using the Gibco-BRL Superscript kit. The RNA was removed by RNAse

digestion, and the cDNA was diluted with TE buffer and adjusted to a final concentration of approximately 70 ng/µl.

The cDNA was generated by the polymerase chain reaction (PCR) procedure, using the Boehringer Mannheim Expand High Fidelity PCR System. Each reaction was performed in a total volume of 50 µl. The components of the reaction were 3 µl (210 ng) template DNA, 200 µM each dNTP, 300 nM each primer as described below, 1X buffer with $MgCl_2$ as supplied by the manufacturer, and 2.6 units of enzyme.

The reaction mixture was assembled with the above components, except for the first strand cDNA template. The first strand cDNA template was added subsequently, and the PCR reaction started as follows:

95°C for 2:00

35 cycles of: 95°C for 0:15
54°C for 0:30
72°C for 1:00

For the first round of PCR, the primers were:

R334N2-L1BAM CGGGATCCCCGCACAAACTTATATGACAACTC (SEQ ID NO:176)

R334N2-R1ECORI CGGAATTCGGTGTCTCATAATGGTAGTGGATAC (SEQ ID NO:177)

The first round of PCR yielded no detectable band as determined by agarose gel electrophoresis, staining with ethidium bromide, and visualization on a UV light box.

A second round of PCR was performed as described above, with the following changes: the template DNA was 0.5 µl of the 1st round PCR reaction, and a different (nested) 3' primer was used. The primers were:

R334N2-L1BAM CGGGATCCCCGCACAAACTTATATGACAACTC (SEQ ID NO:178)

R334N2-R2ECORI CGGAATTCGCAAAAGAGAGGTATAGGGATAAAG (SEQ ID NO:179)

This reaction yielded a strong single band of DNA at approximately 400 bp.

To subclone the PCR product into a plasmid vector for DNA sequencing, we first purified the PCR reaction using the Promega Wizard PCR preps DNA purification system kit, according to the manufacturer's instructions, except the purified DNA was eluted from the column using 25 µl of distilled water. The purified DNA was digested with BamHI and EcoRI and the digested PCR product was isolated by agarose gel electrophoresis on a 1% agarose gel. The DNA product was eluted by electrophoresis into 1% low-melting temperature agarose. The product was purified from the gel by digestion of the low-melting temperature agarose with 5 units of B-agarase I (New England Biolabs) for 1 hour at 40°C in 1X B-agarase buffer provided by the manufacturer, followed by precipitation of the DNA with 1/10 volumes of 3M sodium acetate, pH 5.2 and 2 volumes of isopropanol. Following

incubation of this mixture at -20°C for 30 minutes, the precipitated DNA was recovered by centrifugation at 13,500 X g for 15 minutes, the supernatant was removed, the DNA pellet was air-dried for 10 minutes and resuspended in 10-20 µl of distilled water. We then ligated the PCR product into the plasmid vector pcDNA3.1 (InVitrogen), cut with BamHI and EcoRI and transformed *E. coli*. We screened bacterial colonies for the correct plasmid by preparing mini-prep DNA using the Primm Labs Mini-Prep kit, and analyzed the mini-prep DNA by BamHI and EcoRI restriction digestion and agarose gel electrophoresis.

We sequenced the subcloned PCR products by thermal cycling, using the Big Dye ready reaction mix sequencing kit. For each sequencing reaction, we added: approximately 100-200 ng of mini-prep DNA; 0.8 pmol of sequencing primer; 1 µl 1X BigDye ready reaction buffer (80 mM Tris, 2 mM $MgCl_2$, pH 9.0) and adjusted the volume to 5 µl with distilled water. The following sequencing primers were used:

pcDNA3.1BGH Reverse TAGAAGGCACAGTCGAGG (SEQ ID NO:180)

T7 promoter primer TAATACGACTACTATAGGG (SEQ ID NO:181)

The sequencing reactions were thermal cycled using the following program:

96° for 4:00

25 cycles of: 96°C for 0:30

50°C for 0:15

60°C for 4:00

We purified the cycled DNA by centrifugation through Centriflex gel filtration cartridge spin columns (Edge Biosystems), according to the manufacturer's instructions. The purified DNA was dried by vacuum centrifugation for 30 minutes. The sequencing reactions were analyzed and the DNA sequence determined by gel electrophoresis and fluorescent detection of sequencing products.

ZC334.N3

The first round PCR was performed exactly as ZC334.N2, except the 5' primer contains an HindIII site, and the 3' primer contains and XbaI site, as the Y52A1.N primers. First round primers:

334N3-LIH3 CCCAAGCTTAAAGGCTTAGATGCAGAAAGACC (SEQ ID NO:182)

334N3-RXBA GCTCTAGAGGGATTAAAATCACTCTGTGATTAAG (SEQ ID NO:183)

The first round of PCR yielded no detectable band as determined by agarose gel electrophoresis, staining with ethidium bromide, and visualization on a UV light box.

A second round of PCR was performed as described above; a different (nested) 5' primer was used. The primers were:

334N3-L2H3 CCCAAGCTTTAAAGGTGGACATTGTAGAAGGTTG (SEQ ID NO:184)
334N3-RXBA GCTCTAGAGGGATTAAAATCACTCTGTGATTAAG (SEQ ID NO:185)

This reaction yielded several different sized DNA products, including a strong band of DNA at the predicted size of approximately 350 bp. This 350 bp product was subcloned and sequenced exactly as described for ZC334.N2.

ZC334.N4

The first round PCR was performed exactly as ZC334.N2. Primers contain HindIII and XbaI sites as ZC334.N3. First round primers:
R334N4-LIH3 CCCAAGCTTCCTTCACTTCTCAGCGAAGGAAATG (SEQ ID NO:186)
R334N4-RXBA GCTCAGAGTGCTCATGCTCCGTTATTTGTGC (SEQ ID NO:187)

This reaction yielded a strong single band of DNA at approximately 380 bp after one round of PCR. This product was subcloned and sequenced exactly as described for ZC334.N2.

ZC334.N5

The first round PCR was performed exactly as ZC334.N2. The 5' primer contains a EcoRI restriction site on the 5' end, i.e. CG-GAATTC-N=26; and the 3' primer contains an XhoI site on the end, i.e. CCG-CTCGAG-N=24 for cloning; the HindIII and XbaI sites, which were used as cloning sites for many of the cDNAs, were not used in this case since ZC334.N5 has both internal HindIII and XbaI sites. First round primers:
R334N5-L1ECORI CGGAATTCCTAGAATTTTCACCCCAAATGTTCAG (SEQ ID NO:188)
R334N5-RXHO CCGCTCGAGAAATGTAAGTGATTGGCAAGTTGG (SEQ ID NO:189)

This reaction yielded a strong single band of DNA at approximately 300 bp after one round of PCR. This product was subcloned and sequenced exactly as described for ZC334.N2.

ZC334.N6

The first round PCR was performed exactly as ZC334.N2. Primers contain HindIII and XbaI sites as ZC334.N3. First round primers:
334N6-L1H3 CCCAAGCTTAGAGACTTAGACGCAAAGAGGACC (SEQ ID NO:190)
334N6-RXBA GCTCTAGAGCAGGAAAATTAGCTAAAACATAATG (SEQ ID NO:191)

The first round of PCR yielded no detectable band as determined by agarose gel electrophoresis, staining with ethidium bromide, and visualization on a UV light box.

A second round of PCR was performed using the same two primers that were used in the ZC334.N6 first round reaction, as described above. This reaction yielded several products, including a strong band of DNA at the predicted size of approximately 450 bp.

This 450 bp product was subcloned and sequenced exactly as described for ZC334.N2.

ZC334.N7

The first round PCR was performed exactly as ZC334.N2. The 5' primer contains a EcoRI restriction site on the 5' end, i.e. CG-GAATTC-N=24; and the 3' primer contains an XhoI site on the end, i.e. CCG-CTCGAG-N=25 for cloning; the HindIII and XbaI sites, which were used as cloning sites for many of the cDNAs, were not used in this case since ZC334.N7 has both internal HindIII and XbaI sites. First round primers:

R334N7-L1ECORI CGGAATTCGGCGAAACACTTCCGCCAACTCAC (SEQ ID NO:192)

R334N7-R1XHO CCGCTCGAGACCTACCTCAACTTGGAGGATAAC (SEQ ID NO:193)

The first round of PCR yielded no detectable band as determined by agarose gel electrophoresis, staining with ethidium bromide, and visualization on a UV light box.

A second round of PCR was performed using the same two primers that were used in the ZC334.N7 first round reaction, as described above. This reaction yielded several products, including a band of DNA at the predicted size of approximately 650 bp. This 650 bp product was subcloned and sequenced exactly as described for ZC334.N2.

T10D4.N

The first round PCR was performed exactly as ZC334.N2. Primers contain HindIII and XbaI sites as ZC334.N3. First round primers:

D4N-L2H3 CCCAAGCTTCCTTGCACCTGCCTTCAACCATCAC (SEQ ID NO:194)

D4N-RXBA GCTCTAGATATTCTGACCCCAAAATGACAATC (SEQ ID NO:195)

This reaction yielded a single band of DNA at approximately 700 bp after one round of PCR. This product was subcloned and sequenced exactly as described for ZC334.N2.

T10D4.N2

The first round PCR was performed exactly as ZC334.N2. Primers contain HindIII and XbaI sites as ZC334.N3. First round primers:

RD4N2-L1H3 CCCAAGCTTTTCTGCAGACTTGCAAGGTTAGTTC (SEQ ID NO:196)

RD4N2-R1XBA GCTCTAGAATTCACAAAATAATCAAGACAATC (SEQ ID NO:197)

The first round of PCR yielded no detectable band as determined by agarose gel electrophoresis, staining with ethidium bromide, and visualization on a UV light box.

A second round of PCR was performed using the same two primers that were used in the T1OD4.N2 first round reaction, as described above. This reaction yielded a strong band of DNA at approximately 400 bp. This product was subcloned and sequenced exactly as described for ZC334.N2.

EXAMPLE 2: EXPRESSION ANALYSIS

Analysis of expression patterns of *C. elegans* insulin-like genes was carried out by fusing the transcriptional control regions identified for each gene to a reporter gene encoding green fluorescent protein (GFP), a protein whose expression is easily detected by its fluorescence *in vivo*. Each reporter gene so constructed was then expressed as a transgene in transgenic nematodes. Table 2 entitled "Expression Data" sets forth the

For each *C. elegans* insulin-like gene, putative promoter/enhancer regions were identified in the adjacent genomic sequence (GenBank®, *C. elegans* Genome Project) as regions extending from the predicted start codon of each insulin-like gene to the next gene upstream, identified using the GeneFinder program. If the putative promoter/enhancer region was 6 kilobase pairs (kbp) or less in size, synthetic oligonucleotide primers were designed to amplify the entire region by PCR. For F13B12.N, ZK75.2 and M04D8.1, and the putative promoter/enhancer region was more than 6 kbp or was unbounded (*see* Table 2) by a clearly-defined upstream gene. In these instances, a 2 to 6 kbp segment of upstream region was arbitrarily chosen for amplification, based on available genomic sequence information and favorable primer annealing sites. In addition to the gene-specific sequences incorporated into the PCR primers, each primer also contained restriction enzyme cleavage sites to allow easy insertion into the GFP reporter vector system (pPD117.01): Asc I cleavage sites where incorporated in primers positioned upstream of each enhancer/promoter region, and either Age I or Kpn I sites incorporated into each primer position downstream of the promoter/enhancer. The specific primer pair sequences used to amplify the promoter/enhancer regions of each gene are listed below.

Table 2. Expression Data

Class	Chromosome	Gene	Enhancer (Kb)	Duplicate	Expression Pattern					
					stages	cell bodies in NR (per side)	sensory processes	other processes/ neurons	non-neuronal cells	starved plates
I	IV	F13B12.N	5.2 (not bounded)	independent	embryo-adult	many	amphid	ventral, lateral, tail		same
II	II	ZK75.2	3.7 (not bounded)	independent	embryo (bean), L1 (weak L2, L3?)	many	amphid, IL (2?)	ventral, tail	weak pharynx, vulva	absent/ weak
		ZK75.3	5.7 (bounded)	independent	L1-adult	4?	amphid, IL (2?)	ventral, lateral, dorsal, circumferential		same
		ZK75.1	5.7 (bounded)	same PCR	embryo (2X)-adult	4?	amphid, IL (2?)	ventral, dorsal, tail	hypodermis (L2/3-adult)	same
		ZK84.N2	1.7 (bounded)							
		ZK84.6	3.7 (bounded)	independent	L1-adult	2	amphid, IL (2)	ventral, tail		same
	IV	ZK1251.2	0.6 (bounded)	independent	embryo-adult	4?	amphid, IL (2)	ventral, tail	weak cuticle?	same
		ZK1251.N	1.3 (bounded)	same PCR	embryo-adult	4?	amphid, IL (2)	ventral, tail	vulva, weak cuticle?	same
	X	C06E2.N	3.0 (bounded)	same PCR	L1-adult	2	amphid			same
III	II	C17C3.N	2.3 (bounded)							
		C17C3.4	1.4 (bounded)	independent						
IV	II	ZK84.N	2.1 (bounded							
	III	M04D8.1	3.0 (not bounded)		embryo-adult	2	amphid	ventral, weak tail		same
		M04D8.2	2.2 (bounded)	independent						
		M04D8.3	1.6 (bounded)	independent						
daf-2			3.4 (not bounded)	same PCR	embryo-adult	many	amphid, IL (2)	ventral, lateral, circumferential, vulval, tail	BW muscle, hypodermis (embryo-L4)	same + intestine

List of primers for promoter/enhancer amplification

Gene (PCR product size in kbp)

Sense and antisense primers

F13B12.N (5.2)

TTGGGCGCGCCGTCTTGCATGCAGTTGTCACG (SEQ ID NO:72)

CCAACCGGTATCATTGCGTACTGTCGTAGCGTGTG (SEQ ID NO:73)

ZK75.2 (3.7)

TTGGGCGCGCCTGCTACCGTGGGAATTTTACAAG (SEQ ID NO:74)

CCAACCGGTATCATGGTAGATTTTAGAATGGAAAG (SEQ ID NO:75)

ZK75.3 (5.7)

TTGGGCGCGCCGGAGTTCATCTGGAGGTCACATC (SEQ ID NO:76)

CCAACCGGTATCATTATTCAGAACAGGAATTGATAAATG (SEQ ID NO:77)

ZK75.1 (5.7)

TTGGGCGCCAGATAAATACAGAATGGGCGGAG (SEQ ID NO:78)

CCAACCGGTATCATTCTCTTGGAGCTTTTGAAAAAC (SEQ ID NO:79)

ZK84.N2 (1.7)

TTGGGCGCGCCAGTCGTCCAACAAGCCATCTCC (SEQ ID NO:80)

CCAACCGGTTGCATTTTCCTTGAAGATTGAAG (SEQ ID NO:81)

ZK84.6 (3.7)

TTGGGCGCGCCTAGATTTTCTCCATTCACAAAC (SEQ ID NO:82)

CCAACCGGTATCATTATAATGATATGGATAACGG (SEQ ID NO:83)

ZK1251.2 (0.6)

TTGGGCGCGCCAATCGTTTTCATCATTTTGCTTC (SEQ ID NO:84)

CCAACCGGTATCATCTGGAAAAGTAATATTATAT (SEQ ID NO:85)

ZK1251.N (1.3)

TTGGGCGCGCCTGAAATCTTTATATCCTCTTCAC (SEQ ID NO:86)

CCAACCGGTATCATCTGGAAATAATTAATATCAG (SEQ ID NO:87)

C06E2.N (3.0)

TTGGGCGCGCCTAACACGTGCATTGGAGGCGGAG (SEQ ID NO:88)

CCAACGGTATCATCGTTTCACTCCTCGAATTATTTG (SEQ ID NO:89)

C17C3.N (2.3)

TTGGGCGCGCCATTGGTATCACAAGGATCAAGC (SEQ ID NO:90)

CCAACCGGCATTTTTGTTTTTGGCTGTGATTA (SEQ ID NO:91)

C17C3.4 (1.4)

TTGGGCGCGCCAATTTTGACGACGATCTCCTTC (SEQ ID NO:92)

CCAACCGGTATCATATTTAACGATTCCTACACAAACC (SEQ ID NO:93)

ZK84.N (2.1)

TTGGGCGCGCCGTGTGGAGGTGGTGAATCC (SEQ ID NO:94)

CGGGGTACCCTCATTTCAAAGAAATGTTGAATA (SEQ ID NO:95)

M04D8.1 (3.0)

TTGGGCGCGCCGGAGCCGAACAAGAAAAACCTAC (SEQ ID NO:96)

CCAACCGGTTTCATGGTTCAACTCAAAAAGGAA (SEQ ID NO:97)

M04D8.2 (2.2)

TTGGGCGCGCCAGTTCGTCTCAGCATCATCTTGC (SEQ ID NO:98)

CCAACCGGTTTCATGGTTCAACTCAAAAAGGAA (SEQ ID NO:99)

M04D8.3 (1.6)

TTGGGCGCGCCATGGGATTTTCAGACTCTCAG (SEQ ID NO:100)

CCAACCGGTAACATTATCGAAATTCCAGAAATCCG (SEQ ID NO:101)

The following PCR conditions were used: 95°C for 2 min; either 15 cycles (genomic DNA templates) or 10 cycles (cosmid DNA templates) of the following steps, (1) 95°C for 15 sec, (2) 50°C for 30 sec, and (3) 68°C for a time equivalent to 1 min per kbp of expected product length, and 10 additional cycles with 20 sec added per cycle at step (3). N2 genomic DNA was used as template, except for ZK75.2, ZK75.3, ZK75.1, and ZK84.6, for which cosmid DNA was used. The PCR products were digested with either AscI-AgeI or AscI-KpnI, ligated into similarly-digested PPD117.01 GFP fusion vector, and transformed into *E. coli*. DNA from the resulting clones was prepared using a Qiagen kit, and the correct structure and reading frame of fusion between promoter region and GFP coding region was checked by DNA sequencing.

GFP fusion construct injection

Each GFP fusion construct was injected into wild type worms using a standard protocol for *C. elegans* transformation (*see* Mello et al., 1991, "Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences", EMBO J. 10:3959-3970) at a concentration of 100 µg/ml each GFP fusion plasmid plus 100 µg/ml pRF4 rol-6(d) transformation marker. Stably transformed strains exhibiting a Roller phenotype were established and examined for fluorescence by inspection using an Axioplan microscope (Zeiss). For each GFP fusion construct, two transformant lines which exhibited the highest levels of fluorescence were chosen for further analysis. Duplicate constructs were analyzed for all promoter/enhancer region-GFP fusions, and the patterns of GFP expression were found to be identical for all duplicates (*see* Table 2). Duplicate constructs

were derived from independent PCR reactions for all genes except ZK75.1, ZK1251.N, and C06E2.N.

Structural categories of genes

Comparison of the predicted coding regions of *C. elegans* insulin-like genes reveals a remarkable and unexpected diversity of structures, which are nonetheless clear variations on the common theme that characterizes the insulin superfamily. Structural domains within each predicted *C. elegans* insulin-like protein are annotated in the sequences set forth in FIG. 4 through FIG. 34. In FIG. 3, the sequences of predicted mature forms of the proteins are aligned to one another to highlight features that tend to be conserved compared with the insulin superfamily, as well as to emphasize features that distinguish different Classes of *C. elegans* insulin-like proteins.

We have divided the currently-characterized *C. elegans* insulin-like genes into four Classes based on the protein primary structural characteristics as set forth below.

CLASS I: One *C. elegans* insulin-like gene, F13B12.N has been assigned to Class I. Class I is characterized as having a cleavable C peptide separating the B and A chains. This C peptide possesses processing sites for prohormone convertases, similar to that of vertebrate insulin. Ends generated by proteolytic removal of the C peptide are indicated by the symbols "<<" and ">>" in FIG. 3 for the B and A peptides. Further, Class I is characterized as having an extra pair of Cys residues present which is not found in vertebrate insulins. One Cys residue is located in the B chain and the other Cys residue is located in the A chain. This unique extra pair of Cys residues presumably form an extra inter-chain disulfide bond.

CLASS II: Nine *C. elegans* insulin-like genes, ZK75.1, ZK75.2, ZK75.3, ZK84.6, ZK84.N2, ZK1251.2, ZK1251.N, C06E2.N and TO85G.N have been assigned to Class II. Class II is characterized by the absence of a C peptide. Further, Class II is characterized as having an extra pair of Cys residues.

Still further, Class II is characterized as having a "Pro peptide," which is presumably removed by proteolytic processing from the mature hormone. This Pro peptide is located between the signal sequence and the beginning of the B domain (*i.e.*, similar to the Pro peptide of locust LIRP insulin-like protein). The B and A regions or domains presumably are not cleaved into separate chains in this Class II and the following Classes III-IV.

TO8G5.N is unique in that there is a repositioning of one of the Cys residues in the B domain. In this case, the second Cys residue appears to be moved by four amino acid residues from the end of the presumptive central helix of the B domain towards the middle of the central helix. The repositioning places the Cys residue such that it would project from the same side of the presumptive B domain helix and remain available for disulfide bond formation with the normal partner Cys residue at the end of the second helix of the A domain. Although the spacing of Cys residues in the B domain is unique to insulin-like protein TO8G5.N, it is anticipated that this Cys residue repositioning can be accommodated with relatively small changes in the tertiary structure typical of the insulin superfamily, and no significant changes in secondary structure motifs.

CLASS III: Ten *C. elegans* insulin-like genes, C17C3.4, C17C3.N, C17C3.N2, F41G3.N, F41G3.N2, F56F3.6, Y52A1.N, T28B8.N, T10D4.N and T10D4.N2, have been assigned to Class III. Class III is characterized by the absence of a C peptide. Further, Class III is characterized as having the same number of Cys residues in the B and A domains as found in vertebrate insulin. Some members of this Class lack an intron positioned between the B and A domains within the genomic sequence. FIG. 3 denotes the lack of an intron in this position by the symbol " " at the C-terminus of the B domain and N-terminus of the A domain for C17C3.N2, F41G3.N2, and F56F3.6, and the most N-terminal of the three insulin-like modules of T10D4.N, designated as T10D4.Na, as indicated in FIG. 3.

CLASS IV: Eleven *C. elegans* insulin-like genes, M04D8.1, M04D8.2, M04D8.3, ZK84.N, ZC334.N, ZC334.N2, ZC334.N3, ZC334.N4, ZC334.N5, ZC334.N6 and ZC334.N7, have been assigned to Class IV. Class IV is characterized by the absence of a C peptide. Further, Class IV is characterized as having an extra pair of Cys residues, as in Classes I and II. Still further, Class IV is characterized by the absence of a Cys pair in the A domain; the missing Cys pair in most cases is replaced by hydrophobic residues.

Structural comparison with known genes

With respect to the well-characterized structures of previously-known insulin superfamily proteins, each of the *C. elegans* insulin-like proteins identified herein has at least one novel and significant structural feature which is not typical of the previously-characterized insulin superfamily proteins. These features include: absence of a C peptide; presence of an extra inter-chain Cys pair; absence of a Cys pair in the A chain domain; altered spacing of Cys residues; and/or multiple B domain and A domain pairs in the same

polypeptide. However, these primary structural differences can be accommodated within the overall secondary and tertiary structural framework that is common to the insulin superfamily, as described below.

Peptide domains

Only one of the *C. elegans* insulin-like genes possesses a "connecting" or C peptide between the A and B chain domains (*i.e.*, F13B12.N, Class I). Since the C-terminus of the B chain and the C-terminus of the A chain are relatively close in space within the tertiary structure of insulin, it is quite possible that a continuous main chain could connect presumptive B and A domains without grossly disturbing the overall insulin fold. There is an intriguing aspect of the gene organization of the *C. elegans* insulin-like genes that supports the notion of structural motifs corresponding to the B and A peptides of the insulin superfamily, despite the lack of a C peptide. All *C. elegans* insulin-like genes have introns, and nearly all genes encoding proteins that lack an identifiable C peptide (Classes II through IV) have an intron positioned between the B domain and A domain as indicated in FIG. 3 (the only exceptions are F56F3.6, C17C3.N2, F41G3.N2, and the most N-terminal insulin-like module of T10D4.N indicated as T10D4.Na). Indeed, even the Class I *C. elegans* insulin-like gene, which has a C peptide, also has an intron positioned at the boundary of the B and C peptides. In vertebrates, the most common exon-intron structure of insulin-like genes is that with an intron position either at the boundary or within the C peptide coding region.

One of the *C. elegans* insulin-like genes, T10D4.N, is especially remarkable in terms of domain organization as this gene encodes a single polypeptide which possesses three tandem pairs of B and A domains, or insulin-like "modules", in effect producing a trimeric insulin. Multiple insulin-like modules within the same polypeptide have not been observed previously in any organism. The sequences of the three insulin-like modules within the T10D4.N polypeptide are labeled in FIG. 3 as T10D4.Na, T10D4.Nb, and T10D4.Nc, extending in order from the N-terminus to the C-terminus of the polypeptide. The symbol "-" at the C-terminus of sequences for modules T10D4.Na and T10D4.Nb signifies that the polypeptide sequence continues with the first residue of the sequence in the line below. It is noteworthy that the tandem insulin-like modules in T10D4.N are connected by hydrophobic spacers at the end of the A domain of each module T10D4.Na and T10D4.Nb. Further, the C-terminal module T10D4.Nc contains a tail extending the end of the A domain of the same length and hydrophobic character as the connecting spacer regions. It is also intriguing that immediately adjacent to the T10D4.N gene within genomic DNA is another

insulin-like gene, T1OD4.N2, oriented in the opposite direction which consists of the typical single insulin module. T1OD4.N2 is very closely related in primary sequence to the individual modules that comprise T1OD4.N (see sequence alignments in FIG. 3) and also possesses the tail extending at the end of the A domain that is similar in size and character to the tail and connecting spacers in the trimeric T1OD4.N.

CYS Residues

Most *C. elegans* insulin-like proteins possess an extra pair of Cys residues (Classes I, II and IV) and it is striking that there is a consistent spatial positioning of them (*see* the alignment of FIG. 3). One extra Cys is found toward the C-terminal end of the B chain (*i.e.*, B region or domain) and the other extra Cys is found toward the C-terminal end of the A chain (*i.e.*, A region or domain). These two positions are expected to be very close in space within the known tertiary structure of insulin superfamily proteins. Thus, it is quite possible that the extra Cys residues in the *C. elegans* insulin-like proteins form a disulfide bond that further stabilizes the structure. This situation is reminiscent of that previously noted for extra Cys residues within the MIP family of insulin-like proteins from freshwater snail. However, in the case of the MIP proteins, the extra Cys residues are positioned at the N-terminal regions of the A and B chains (*see* FIG. 2).

Some *C. elegans* insulin-like proteins (*i.e.* Class IV) are missing a pair of Cys residues in the A domain that are invariably found in the previously-characterized insulin superfamily members and which form an intra-chain disulfide bond that stabilizes a bend in the A chain structure. It is notable that, in many of the *C. elegans* Class IV proteins, there appears to be a concerted replacement of these two Cys residues with either aromatic or aliphatic residues. Such substitutions are consistent with the normal placement of this disulfide linkage within the hydrophobic core between the A and B chains. It seems that in these *C. elegans* Class IV insulin-like proteins, a strong covalent linkage has been substituted with a weaker stacking or hydrophobic interaction between side chains in these positions. It is relevant that all *C. elegans* insulin-like proteins that are "missing" a pair of Cys residues within the A domain also have an "extra pair" of Cys residues at the ends of the B and A domains, as described above.

Several *C. elegans* insulin-like proteins are highly unusual by virtue of having an abnormal spacing between conserved Cys positions (T08G5.N, Y52A1.N, F56F3.6, T28B8.N, T10D4.N, T10D4.N2 and ZC334.N. *see* FIG. 3). Nonetheless, as indicated in the sequence alignment of FIG. 3, the changes in spacing can be viewed as relatively small alterations which are not expected to cause large-scale changes in structure that would

deviate from that typical of the insulin superfamily. The "repositioning" of one Cys residue within the B domain of T08G5.N was discussed previously. For other insulin-like genes with altered spacing of Cys residues, the changes in spacing can be viewed as small insertions or deletions within structural transitions of the typical insulin fold. Thus, Y52A1.N can be viewed as having a deletion of three residues (symbolized by "---" in FIG. 3) that shortens the loop connecting the two helices of the A domain. Conversely, ZC334.N and insulin-like modules T1OD4.Nb and T1OD4.Nc of T1OD4.N can all be viewed as having an insertion of a dipeptide of either "Ser Gly", "Pro Glu", or "Ser Ala", respectively, within the loop connecting the two helices of the A domain. Also, T1OD4.N2 and modules T1OD4.Na, T1OD4.Nb, and T10D4.Nc of T1OD4.N can each be viewed as having an insertion of a single residue, either "Ile", "Phe", "Val", or "Val", respectively, at the end of the second helix of the A domain. Finally, F56F3.6 and T28B8.N can be viewed as having an insertion of a tripeptide having the sequence "Pro Pro Gly" within the turn that immediately precedes central helix of the B domain. It is particularly intriguing that the presence of both insertions and deletions of this sort within the *C. elegans* insulin-like proteins points to an ability to accommodate more variation within the insulin protein structure than had been appreciated from sequences of previously described insulin superfamily proteins.

EXAMPLE 3: GENERATION AND GENETIC ANALYSIS OF NEMATODES WITH ALTERED INSULIN-LIKE GENES

C. elegans insulin-like genes are important tools for creating genetically-engineered nematodes. Genetically-engineered nematodes may harbor: (a) deletions or insertions in an insulin-like gene or genes; (b) interfering RNAs derived from such genes; (c) and/or transgenes for mis-expression of wild-type or mutant forms of such genes. Such *C. elegans* strains with laboratory-generated alterations in insulin-like genes are useful for many purposes. Examples of such purposes include: (a) identification of insulin-like genes that participate in biochemical and/or genetic pathways that constitute possible pesticide targets, as judged by phenotypes such as non-viability, block of normal development, defective feeding, defective movement, or defective reproduction; (b) identification of insulin-like genes that participate in genetic and/or biochemical pathways that relate to therapeutic applications associated with the insulin superfamily hormones, such as metabolic control, growth regulation, differentiation, reproduction, and aging, through the generation of phenotypes associated with those functions in the altered *C. elegans* strains; and (c) as substrates for large-scale genetic modifier screens aimed at systematic identification of other

components of these genetic and/or biochemical pathways that serve as novel drug targets, diagnostics, prognostics, therapeutic proteins, pesticide targets or protein pesticides.

Methods for creation and analysis of *C. elegans* strains having modified expression of insulin-like genes are described below. Expression modification methods include any method known to one skilled in the art. Specific examples include but are not limited to EMS chemical mutagenesis, Tc1 transposon mutagenesis, double-stranded RNA interference, and transgene-mediated mis-expression. In the creation of transgenic animals, it is preferred that heterologous (*i.e.*, non-native) promoters be used to drive transgene expression.

EXAMPLE 4: EMS CHEMICAL DELETION MUTAGENESIS

Ethyl methanesulfonate (EMS) is a commonly-used chemical mutagen for creating loss-of-function mutations in genes-of-interest in *C. elegans*. Approximately 13% of mutations induced by EMS are small deletions. With the methods described herein, there is approximately a 95% probability of identifying a deletion-of-interest by screening 4 x 10^6 EMS-mutagenized genomes. Briefly, this procedure involves creating a library of several million mutagenized *C. elegans* which are distributed in small pools in 96-well plates, each pool composed of approximately 400 haploid genomes. A portion of each pool is used to generate a corresponding library of genomic DNA derived from the mutagenized nematodes. The DNA library is screened with a PCR assay to identify pools that carry genomes with deletions-of-interest, and mutant worms carrying the desired deletions are recovered from the corresponding pools of the mutagenized animals. Although EMS is a preferred mutagen to generate deletions, other mutagens can be used that also provide a significant yield of deletions, such as X-rays, gamma-rays, diepoxybutane, formaldehyde and trimethylpsoralen with ultraviolet light.

Nematodes may be mutagenized with EMS using any procedure known to one skilled in the art, such as the procedure described by Sulston and Hodgkin (1988, Methods, pp. 587-606, *in* The nematode *Caenorhabditis elegans*, Wood, Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). Following exposure to the mutagen, nematodes are dispensed into petri dishes, incubated one to two days, and embryos isolated by hypochlorite treatment (*Id.*) Embryos are allowed to hatch and L1 larvae are collected following overnight incubation. The larvae are distributed in petri plates at an average density of 200 animals per plate and incubated for 5 to 7 days until just starved. A sample of nematodes is collected from each plate by washing with a solution of distilled water, and the nematodes washed from each plate are placed in one well of a 96-well plate. Worms are

lysed by addition of an equal volume of lysis buffer (100 mM KCl, 20 mM Tris-HCl pH 8.3, 5 mM $MgCl_2$, 0.9% Nonidet P-40, 0.9% Tween-20, 0.02% gelatin, and 400 µg/ml proteinase K) followed by incubation at -80°C for 15 minutes, 60°C for 3 hours, and 95°C for 15-30 minutes. The DNA-containing lysates are kept by storage of plates at -80°C until analyzed further. Live nematodes from each plate are aliquoted into tubes within racks for storage at -80°C, such that the physical arrangement of tubes of live animals is the same as the arrangement of corresponding DNA lysates in the 96-well plates.

A pooling strategy is used to allow efficient PCR screening of the DNA lysates. The pools are made from each 96-well plate by mixing 10 µl of lysate from 8 wells comprising each column of wells in a plate. The pooled lysates for each column are used for screening with PCR. PCR primers are designed for each locus-of-interest to be about 1.5 to 12 kb apart, depending on the size of the locus, such that deletions encompassing the entire coding regions of insulin-like genes can be detected following a previously-described procedure (*see* Plasterk, 1995, Reverse genetics: from gene sequence to mutant worm, Methods in Cell Biology 48:59-80). For each region, two sets of primer pairs are chosen for carrying out a nested PCR strategy such that an outside set is used for the first round of PCR and an inside set is used for the second round of PCR. The second round of PCR is performed to achieve greater specificity in the reaction.

The first round PCR reactions are performed in duplicate for each pool with reactions carried out in a 96-well plate. Each reaction contains 18 µl of the following mixture and 2 µl of each pooled lysate:

reaction buffer provided by the manufacturer (*e.g.*, Boehringer Mannheim Biochemicals)
2.5 mM $MgCl_2$
0.2 mM each dNTP
0.5 µM each gene-specific primer
1.7 units Expand Hi Fidelity enzyme mix (Boehringer Mannheim Biochemicals)
to 18 µl per reaction with dH_2O

The reactions are carried out using the same general temperature cycling parameters except that the extension time is varied depending on the normal distance between the primer pairs as follows:

4 kb wild-type product or shorter: 1 minute extension time
4-6 kb wild-type product: 2 minute extension time
6-12 kb wild-type product: 4 minute extension time

The temperature cycling conditions used are 94°C for 3 minutes, then 35 cycles of the following: 94°C for 40 seconds, 55°C for 1 minute, and 72°C for the number of minutes of extension time described above.

The second round of PCR is performed essentially as above, except that 15 µl of mixture containing the following was aliquoted to each reaction:

reaction buffer provided by the manufacturer
1.5 mM $MgCl_2$
0.2 mM each dNTP
0.5 µM each gene-specific primer
1.7 units of Expand Hi fidelity enzyme mix
to 15 µl per reaction with dH_2O

A small amount of first-round reaction products is transferred to the second-round reaction mixtures using a 96-pin replicator. The same temperature cycling sequence is used for the second round as described for the first round.

Products of the second round of PCR may be analyzed by electrophoresis in 1% agarose gels. If a potential deletion product is observed in at least one of the two reactions, two rounds of PCR are performed as described above on lysates from each individual well derived from the column corresponding to the positive pool. This results in the identification of a positive "address," *i.e.*, a specific well within an individual plate, containing a deletion mutant. The positive address is re-tested in quadruplicate using two rounds of PCR as described above, and the product is gel purified and sequenced directly to confirm the presence of the desired deletion.

For example, two deletions have been identified and characterized by DNA sequencing, using the procedures described above, that remove the *C. elegans* insulin-like gene ZK75.1.

Once a positive address has been identified and confirmed by sequence analysis, approximately 300 individual worms from the relevant plate are cloned onto separate, fresh plates. When F1 animals are present on the plate, the parent nematodes are placed into buffer and lysed as described above. The same primer pairs and cycling conditions used to identify the deletion are used to perform PCR on these animals. Once a single animal carrying the deletion has been identified, its progeny are cloned and examined using the same conditions described above, until a homozygous population of deletion animals is obtained.

Detailed protocols which may be used for EMS mutagenesis of the genes identified herein are set forth below.

Mutagenizing nematodes

Plates crowded with L4 hermaphrodite worms are washed off with M9 buffer into 15 ml tubes and centrifuged. The worms are washed 2X with M9 buffer and resuspended in 9 ml of M9 buffer and transferred to a 50 ml tube.

In a chemical fume hood, 1 ml of M9 buffer and 62 μl of EMS are added to a microfuge tube. Close tube and shake to mix M9 and EMS. The EMS/M9 mixture is then added to the 9 ml of worms. This is a concentration of 50 mM EMS in 10 ml of worms in suspension. Rotate suspension on a rotation device (e.g., Nutator) for 4 hours. After the incubation, wash worms with M9 buffer 3X.

Plate animals to plates with thick lawns of bacteria and place them at 20 °C for about 24 hours until they become full of eggs as adults. Hypochlorite treat worms to kill adults and isolate embryos (*see* below).

Isolating worm embryos

The following protocol may be used to isolate mutagenized worm embryos following the above EMS chemical treatment:

1. wash worms off plates into a 15 ml tube in a total of 15 ml sterile water
2. spin down worms 30 sec at about 15K rpm and wash 2X in water
3. rinse worms briefly in 4 ml hypochlorite solution (6.6 ml water, 400 µl 5 M KOH, 1 ml 10% Na hypochlorite) and spin down
5. add remaining 4 ml hypochlorite solution and transfer a drop to a watch glass to observe the reaction under a dissecting microscope
6. as soon as adults start to burst at vulva and release embryos, adults are broken open by passage through a 21 gauge needle 2-3X
7. quickly fill tube with M9 buffer and spin down eggs
8. rinse 3X with M9 buffer
9. filter embryos through 52 µm mesh in 30 ml M9 into a 50 ml tube (if volume of embryos < 0.5 ml, embryos are resuspend in 8 ml M9 buffer in a 15 ml tube)
10. rotate embryos on nutator at 15°C overnight
11. spin down L1 larvae and plate on 3-8 large NGM plates seeded with concentrated *E. coli*

A typical library may contain 6668 lysates representing 2.18 million haploid genomes.

List of primers for EMS analysis (EMS table)

Genes screened - (product size)	primer name	primer sequence
C06E2.N (X) - (2.1 kb)	CO6E2-1 (round 1 forward)	CAAACAGTTGTAGCTCAAAGGC (SEQ ID NO:104)
	CO6E2-4 (round 1 reverse)	GCATACGGTACCTATTCGTTTC (SEQ ID NO:105)
	CO6E2-2 (round 2 forward)	AGCTCAAAGGCCAAATGTGTG (SEQ ID NO:106)
	CO6E2-3 (round 2 reverse)	AACAAACCCTACAGTTACTGGG (SEQ ID NO:107)
ZK75.2/75.3(II) -(3.6 kb)	ZK75-31 (round 1 forward)	GCTATCCACCTGTCCAACCTAC (SEQ ID NO:108)
	ZK75-35 (round 1 reverse)	GGAGGCTCTTTACTCGCCTTAC (SEQ ID NO:109)
	ZK75-32 (round 2 forward)	TACAGGCTGTCCTTCTGTTACG (SEQ ID NO:110)
	ZK75-34 (round 2 reverse)	TCCACTATTCCGGTAATACCTC (SEQ ID NO:111)
ZK1251.N/ZK1251.2 (IV) - (3.5kb)	ZK1251-W1 (round 1 forward)	GTAAGAAATCGAGAGTCACGCC (SEQ ID NO:112)
	ZK1251-W4 (round 1 reverse)	GTCTTCACTATCAAACGGGAGG (SEQ ID NO:113)
	ZK1251-W2 (round 2 forward)	CTGCCTCAAGGAGGAGTTACAC (SEQ ID NO:114)
	ZK1251-W3 (round 2 reverse)	ATTTATCCCCACGTGAGAGAGG (SEQ ID NO:115)
ZK75.2/.3/.1/84.N2/84.6 (II) - (12.7 kb)	ZK75-31 (round 1 forward)	see above
	ZK75-W4 (round 1 reverse)	CACTGGGATGACAGATTTGATG (SEQ ID NO:116)
	ZK75-32 (round 2 forward)	see above
	ZK84-3B (round 2 reverse)	TGATGAGACACGGGTGAAACG (SEQ ID NO:117)

ZK75.1/84.N2/84.6 (II) - (4.7 kb)	ZK75-1F (round 1 forward)	GAACGGATAAAAAGGCGGAGC (SEQ ID NO:118)
	ZK75-W4 (round 1 reverse)	see above
	ZK75-2A (round 2 forward)	TTGATGTGACCTCCAGATGAAC (SEQ ID NO:119)
	ZK84-3B (round 2 reverse)	see above
M04D8.1/.2/.3 (III) - (5 kb)	M04D8-1 (round 1 forward)	GCAGCACACTCTTGTTTTCAGC (SEQ ID NO:120)
	M04D8-4 (round 1 reverse)	CAAATCACTCACTITCCTGCG (SEQ ID NO:121)
	M04D8-2 (round 2 forward)	TTCAAGTGTCCTTGTATCCGTG (SEQ ID NO:122)
	M04D8-3 (round 2 reverse)	GCATAGAATGGCGGAAGAT CAC (SEQ ID NO:123)
F13B12.N (IV) - (2.1 kb)	F13B12-1 (round 1 forward)	CTTCCAAATTTGTCCTGACTGC (SEQ ID NO:124)
	F13B12-4 (round 1 reverse)	AATTGCAGGAGTCGAAGTTTCC (SEQ ID NO:125)
	F13B12-2 (round 2 forward)	AACGAGCAGACAGGAAATC ATC (SEQ ID NO:126)
	F13B12-3 (round 2 reverse)	TGTGACAGCATGTTTGAACGTC (SEQ ID NO:127)
ZK75.1 (II) - (3.7 kb)	ZK75-11 (round 1 forward)	AGTTGTCAAGAAGTGCGTCAAG (SEQ ID NO:128)
	ZK75-1B (round 1 reverse)	GAGATGGCTTGTTGGACGAC (SEQ ID NO:129)
	ZK75-12 (round 2 forward)	GACAAAATCACGTCACGAAGT (SEQ ID NO:130)
	ZK75-13 (round 2 reverse)	TTACTTTTCTGGGCAGCAAGC (SEQ ID NO:131)

Results of an example EMS screen

The following results were obtained in an example EMS screen.

C06E2.N region: 2.3 million haploid genomes screened

ZK75.2/.3 region: 1.2 million haploid genomes screened

ZK1251.2/.N region: 1.2 million haploid genomes screened

ZK75.1 region: 800,000 haploid genomes screened

Two confirmed deletions have been obtained in the ZK75.1 region, as follows:

(1) ZK75.1Δ1 deletes nucleotides 15,182-17,369 of cosmid ZK75.1

(2) ZK75.1Δ2 deletes nucleotides 15,430-17,879 of cosmid ZK75.1

ZK75.2/.3/.1/84.N2/84.6 region: 875,000 haploid genomes screened

ZK75.1/84.N2/84.6 region: 2.1 million haploid genomes screened

M04D8.1/.2/.3 region: 460,000 haploid genomes screened

F13B12.N region: 1.9 million haploid genomes screened

EXAMPLE 5: Tc1 TRANSPOSON INSERTION MUTAGENESIS

The transposable element Tc1 may also be used as a mutagen in *C. elegans* since insertion of the transposable element into a gene-of-interest can result in the inactivation of gene function. Starting with a strain that contains a high copy number of the Tc1 transposable element in a mutator background (*i.e.*, a strain in which the transposable element is highly mobile), a Tc1 library containing approximately 3,000 individual cultures is created as previously described (*Id.*). The library is screened for Tc1 insertions in the region of interest using the polymerase chain reaction with one set of primers specific for Tc1 sequence and one set of gene-specific primers. Because Tc1 exhibits a preference for insertion within introns, it is sometimes necessary to carry out a secondary screen of populations of insertion animals for imprecise excision of the transposable element, which can result in deletion of part or all of the gene of interest (generally, 1-2 kb of genomic sequence is deleted). The screen for Tc1 deletions is performed and deletion animals are recovered in the same manner as for the EMS screen described above.

Using such procedures, *C. elegans* strains have been isolated that contain Tc1 transposon insertions within or neighboring the following insulin-like genes: ZK1251.1/ZK1251.N, CO6E2.N, and F13B12.N. Detailed methods are set below.

Tc1 library construction

A Tc1 transposon insertion library was constructed according to published protocols by Zwaal et al., 1993, Proc. Natl. Acad. Sci. U.S.A. 90:7431-7435; and Plasterk, 1995, Reverse Genetics: From Gene Sequence to Mutant Worm, *in Caenorhabditis elegans*: Modern Biological Analysis of an Organism (Epstein and Shakes, Eds.) pp. 59-80.

Size of typical library: 3 sets of 960 cultures
Analysis of library: By sets of 960 cultures
Dimensions of set: 10 racks of 8 X 12 as follows:
Row (8): A-H
Column (12): 1-12
Plate (10): p1-p10

Culturing worms

POUR 100-mm NGM (2X peptone) plates--2880 plates total
SEED with *E. coli* in sterile hood
CULTURE 5-10 non-synchronized *mut-2* (MT3126) animals per plate--250 plates/day for 12 days:

- PREPARE suspension of MT3126 in M9 buffer in dish
- TRANSFER 5 μl of suspension onto plates
- COUNT # worms on first few plates
- INCUBATE @20°C for 11-12 days

ADD 4 ml M9 buffer to plate
SHAKE plates O/N @18-20°C

Storage of worms

PREPARE Costar racks (3 racks required per 96 cultures)--90 racks total:

- MARK racks clearly on front, side, and top
- MARK individual tubes in each rack

ALIQUOT each culture into 3 racks (8 X 12)--240 cultures/day for 12 days:

- ADD few drops of fresh M9 buffer if <1 ml suspension on plate
- TRANSFER 400 μl suspension to identical positions on 2 racks (for freezing) and remaining suspension to identical position on 3RD RACK (for DNA analysis)

FREEZE 2 racks for survival:

- ADD 400 μl freezing solution to each tube:

30% glycerol (v/v)
25 mM KPO_4, pH 6.6
50 mM NaCl
2.5 μg/ml cholesterol

- CLOSE tubes with sterile caps (8 caps on a strip, Costar)
- COVER rack with lid
- MIX M9 buffer and freezing solution by inverting rack several times
- WRAP racks in cotton wool and 2 towels for slow freezing O/N @-80°C
- UNWRAP racks and store in separate freezers @ -80°C

Lysate preparation (3rd rack)

REMOVE M9 buffer supernatant from sedimented worm suspension
WASH 1X with cold H_20--960 cultures/day for 3 days
CENTRIFUGE for 3 minutes to pellet worms and ice for 30 sec
REMOVE supernatant
(FREEZE worm pellets or LYSE directly)
ADD 200 μl Cell Lysis Solution (Gentra Kit) and 2 μl Proteinase K (10 mg/ml) to each pellet
CLOSE tubes with sterile caps (8 caps on a strip, Costar)
COVER rack with lid
INCUBATE @ 55°C for 3 hrs - O/N (invert. occasionally)
STORE @ -20 or -80°C

DNA preparation

POOL lysates in 3-D matrix: Pool Rows (individual A - H by plate)
240 pools total
8 pools/plate
12 lysates/pool
pool = 240 μl
TRANSFER 20 μl of each lysate/row to a pool--80 pools/day for 3 days
VORTEX

1-D Address: Row	2-D Address: Plate
Pool Rows (cumulative A - H)	240 pools total
24 pools total	8 pools/plate
10 mixed lysates/pool	12 lysates/pool
120 lysates(total)/pool	pool = 60 µl
pool = 1.8 ml (180 µl of each mixed lysate)	

- TRANSFER 180 µl of each mixed lysate/row to a pool
- PURIFY DNA by Gentra kit--24 DNA preps
- RESUSPEND in TE: 10 mM Tris-HCl 1 mM EDTA, pH 7.6
- STORE @ -20°C

88 DNA preps/day for 3 days

(This stock may be used for many searches: 1OX-50X dilutions used.)

Library screening

A library is screened in individual Tiers, each library having three Tiers. Each Tier is composed of 1,000 lysates or 200,000 haploid genomes. Lysates are pooled according to above references. First dimension screen involves PCR on 8 samples of pooled DNA from 10 96-well plates. Second dimension screen determines on which of the 10 96-well plates the mutant resides (involves screening of 10 DNA pools). Third dimension screen determines the "address" of a particular mutant (*i.e.*, in which column and row a particular mutant resides - via screening of 12 individual lysates from a single row). First dimension reactions are done in quadruplicate; second and third are done in triplicate.

Two rounds of PCR are performed; PCR is performed with a pair of gene-specific primers and a pair of Tc1-specific primers. Two different pairs of Tc1 primers are used: one pair points outward from the left of the transposon, and the other pair points outward from the right (these primer pairs are described in the references cited above).

The first and second round PCR for each dimension is performed in 15 µl using the following in each reaction:

1X PCR buffer provided by the manufacturer (Perkin Elmer)

1.5 mM $MgCl_2$

0.2 mM dNTPs

0.5 µM of the Tc1 and the gene-specific primer

0.5 units of Perkin Elmer Taq Polymerase

H_2O to 13 μl for the first round reactions, and to 15 μl for the second round

First and Second dimension: 2 μl of 1:20 DNA is added; 1:10 DNA is added to the third dimension reactions. A small amount of first round reaction is transferred to the second round using a pin replicator. PCR cycling conditions are: 94 for 3 minutes; then 94 for 40 seconds, 58 for 1 minute, 72 for 2 minutes for 35 cycles; then 72 for 2 minutes.

LIST OF PRIMERS FOR TC1 ANALYSIS (TC1

Genes screened	Oligo name	Oligo sequence
All	*Tc1 L1 (round 1 left)	CGTGGGTATTCCTTGTTCGAAG CCAGCTAC (SEQ ID NO:132)
	*Tc1 L2 (round 2 left)	TCAAGTCAAATGGATGCTTGAGA (SEQ ID NO:133)
	*Tc1 R1 (round 1 right)	TCACAAGCTGATCGACTCGATG CCACGTCG (SEQ ID NO:134)
	*Tc1 R2 (round 2 right)	GATTTTGTGAACACTGTGGTGAAGT (SEQ ID NO:135)
ZK75.2/.3/.1/84.N2/84.6	ZK75-31 (round 1)	SEE EMS TABLE
	ZK75-32 (round 2)	SEE EMS TABLE
	ZK75-35 (round 1)	SEE EMS TABLE
	ZK75-34 (round 2)	SEE EMS TABLE
	ZK75-1F (round 1)	SEE EMS TABLE
	ZK75-2A (round 2)	SEE EMS TABLE
	ZK75-W4 (round 1)	SEE EMS TABLE
	ZK84-3B (round 2)	SEE EMS TABLE
	ZK75-M4 (round 1)	TTATTACATCCGTCACTGCGTC (SEQ ID NO:136)
	ZK75-M3 (round 2)	GCGTCCTTATTCAGAATTCCAG (SEQ ID NO:137)
ZK1251.N/ZK1251.2 (IV)	ZK1251-W4 (round 1)	SEE EMS TABLE
	ZK1251-W3 (round 2)	SEE EMS TABLE
	ZK1251-24 (round 1)	CTTGTGACTTCAAGCCCACTTC (SEQ ID NO:138)
	ZK1251-23 (round 2)	GGTTATGAACCGATTAGGCTCC (SEQ ID NO:139)
	ZK1251-N1 (round 1)	GTAGCCTTCCGGGGTTAAAATC (SEQ ID NO:140)
	ZK1251-N2 (round 2)	GATCTCGCGCTATGTTTTGAG (SEQ ID NO:141)

C0632.N (X)	C06E2-1A (round 1)	GACAGCTGAAGCTGACCAAAC (SEQ ID NO:142)
	C06E2-2A (round 2)	CAGGAGTTAAACGTGGTCACTG (SEQ ID NO:143)
	C06E2-4 (round 1)	SEE EMS TABLE
F13B12.N (IV)	F13B12-1 (round 1)	SEE EMS TABLE
	F13B12-2 (round 2)	SEE EMS TABLE
	F13B12-4 (round 1)	SEE EMS TABLE
	F13B12-3 (round 2)	SEE EMS TABLE
M04D8.1/.2/.3 (III)	M04D8-1 (round 1)	SEE EMS TABLE
	M04D8-4 (round 1)	SEE EMS TABLE
	M04D8-2 (round 2)	SEE EMS TABLE
	M04D8-3 (round 2)	SEE EMS TABLE

Results of tc1 screen

Five confirmed Tc1 insertions have been found in or near the following *C. elegans insulin-like genes*: one insertion near ZK1251.2/.N; two insertions near C06E2.N; and two insertions in F13B12.N.

EXAMPLE 6: DOUBLE-STRANDED RNA INTERFERENCE ANALYSIS

The function of the *C. elegans* insulin-like genes identified herein may be characterized and/or determined using a method based on the interfering properties of double-stranded RNAs derived from the coding regions of the identified genes (*see* Fire et al., 1998, Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, Nature 391:806-811). In this method, sense and antisense RNAs derived from a substantial portion of a *C. elegans* insulin-like gene are synthesized *in vitro* from phagemid DNA templates containing cDNA clones of insulin-like genes which are inserted between opposing promoters for T3 and T7 phage RNA polymerases, or from PCR products amplified from coding regions of insulin-like genes, where the primers used for the PCR reactions are modified by the addition of phage T3 and T7 promoters. The resulting sense and antisense RNAs are annealed in an injection buffer and the double-stranded RNA injected into *C. elegans* hermaphrodites. Progeny of the injected hermaphrodites are inspected for phenotypes-of-interest. Other methods can also been employed for generating mutant phenotypes in nematodes using single-stranded antisense DNA or RNA species, as described above. However, single-stranded methods may be less effective in nematodes than that of double-stranded RNA interference (*see* Guo and Kemphues, 1995, *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed, Cell 81:611-620; *see also* Fire, 1991, Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle, Development 113:503-514).

EXAMPLE 7: MIS-EXPRESSION ANALYSIS

Mis-expression (*i.e.*, ectopic expression, abnormal expression) of wild-type and/or mutant *C. elegans* insulin-likegenes so as to create transgenic animals is another useful method for the analysis of gene function in nematodes (Mello and Fire, 1995, DNA transformation, Methods in Cell Biology 48:451-482). Such transgenic animals may be created to contain gene fusions of the coding regions of insulin-like genes joined (*i.e.*, operably linked) to a specific promoter whose regulation has been well characterized. Such

a specific promoter may be used as a heterologous promoter (*i.e.*, a promoter which is not naturally linked to the gene). Examples of promoters that can be used to drive such mis-expression of insulin-like genes include but are not limited to: the heat shock gene promoters *hsp 16-2* and *hsp 16-41*, useful for temperature-induced expression; the *myo-2* gene promoter, useful for pharyngeal muscle-specific expression; the *hlh-1* gene promoter, useful for body-muscle-specific expression; and the *mec-3* gene promoter, useful for touch-neuron-specific gene expression. Gene fusions for directing the mis-expression of insulin-like genes are incorporated into a transformation vector which is injected into nematodes along with a plasmid containing a dominant selectable marker, such as *rol-6*. Transgenic animals are identified as those exhibiting a roller phenotype, and the transgenic animals are inspected for additional phenotypes of interest created by mis-expression of the insulin-like gene.

EXAMPLE 8: ANALYSIS OF MUTANT PHENOTYPES

After isolation of nematodes carrying mutated or mis-expressed insulin-like genes, or inhibitory RNAs, animals are carefully examined for phenotypes-of-interest. For the situations involving deletions or Tc1 insertions in insulin-like genes, nematodes are generated that are homozygous and heterozygous for the mutant insulin-like genes.

Examples of specific phenotypes that may be investigated include but are not limited to: lethality, sterility, reduction in brood size, egg-laying defects, dauer constitutive, dauer defective, increased life span, decreased life span, defective locomotion, defective chemotaxis, defective thermotaxis, abnormal body shape, abnormal body size, and alterations in the morphogenesis of specific organs, such as the vulva, nervous system, gut, or musculature (*see* Hodgkin, 1997, Appendix I: Genetics, pp. 882-1047, *in C. elegans* II, Riddle et al., Eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

EXAMPLE 9: ANALYSIS OF GENETIC INTERACTIONS AND MULTIPLE MUTANTS

Another approach that may be used to probe the biological function of the insulin-like genes identified herein is by using tests for genetic interactions with other genes that may participate in the same, related, interacting, or modifying genetic or biochemical pathways. In particular, since it is evident that there are closely-linked clusters of insulin-like genes in the *C. elegans* genome, this raises the possibility of functional redundancy of one or more genes. Consequently, it is of interest to investigate the phenotypes of nematodes containing mutations (such as deletions or Tc1 insertions as described above)

that knock-out the function of more than one insulin-like gene. Such strains carrying mutations in multiple genes can be generated by cross breeding animals carrying the individual mutations, followed by selection of progeny that carry the desired multiple mutations. Alternatively, multiple insulin-like genes can be inactivated by the simultaneous injection of double-stranded RNAs derived from each gene using the method of double-stranded RNA interference described above.

One specific question-of-interest is genetic analysis of interactions of insulin-like genes with other well-characterized *C. elegans* genes and pathways. Thus, double mutant nematodes may be constructed that carry mutations in an insulin-like gene and another gene-of-interest. It is of particular interest to test the interaction of the insulin-like genes with other genes involved in the dauer formation and life span pathway, especially those that exhibit homology to insulin signaling components in vertebrates. For example, nematodes carrying mutations in insulin-like genes and either a loss-of-function mutation of *daf-16*, a hypomorphic allele of *daf-2*, a hypomorphic allele of *age-1*, would be of use in investigating the involvement of different insulin-like genes in the dauer formation and life span pathways. Also, transgenic animals mis-expressing insulin-like genes which further carry mutations in *daf-2* are of interest, *e.g.*, for examining genetic interactions between the insulin-like genes and the dauer formation and life span pathways. Other genetic interactions may be tested based on the phenotypes observed for alterations of the insulin-like genes alone. For example, if alteration of insulin-like genes produces an abnormal body size, mutations in these insulin-like genes could be tested for interactions with other genes that also affect body size, such as *daf-4*, *sma-2* and *sma-3*.

EXAMPLE 10: GENETIC MODIFIER SCREENS

The initial characterization of phenotypes created by mutations in single or multiple insulin-like genes is expected to lead to the identification of nematode strains that exhibit phenotypes appropriate for large-scale genetic modifier screens aimed at discovering other components of the same pathway. For example, it is of particular interest to identify those insulin-like genes that encode ligands of the *daf-2* receptor. Potential *daf-2* ligands (agonists) might be revealed by the genetic interaction analysis described above as those insulin-like genes which, when mutated alone or in combination, exhibit the following properties: (a) a dauer constitutive phenotype similar to that observed in *daf-2* mutant animals; and (b) suppression of the dauer constitutive phenotype when insulin-like gene mutations are tested in combination with mutations in the *daf-16* gene (an antagonist of the pathway). There are, however, many other phenotypes that could be suitable starting points

for large-scale genetic modifier screens, including a defective egg-laying phenotype, an abnormal lipid accumulation phenotype (*e.g.*, as revealed by staining with lipid-specific dyes), and decreased or increased life span phenotypes.

The procedures involved in a typical genetic modifier screen are described below (*see also* Huang and Sternberg, 1995, Genetic discussion of developmental pathways, Methods in Cell Biology 48:97-122). In general, hermaphrodites carrying mutations in insulin-like genes are exposed to a mutagen, such as EMS or trimethylpsoralen with ultraviolet radiation. The descendants of such animals are then screened for the rare individuals that display suppressed or enhanced versions of the original phenotype, and any new mutations detected are presumed to alter other genes that participate in the same phenotype-generating pathway. In a pilot-scale genetic screen, 10,000 or fewer mutagenized nematodes would be inspected; in a moderate-scale genetic screen, about 30,000 to 100,000 mutagenized animals would be inspected; and in a large-scale genetic screen, more than 100,000 mutagenized animals would be inspected.

Next, nematodes identified with suppressor or enhancer mutations are isolated, and populations of descendants of these animals are expanded. The newly-identified "modifier" genes that are altered by these suppressor or enhancer mutations are mapped using a combination of genetic and molecular methods. Such newly-identified modifier mutations may also be isolated away from the mutations in the insulin-like genes by genetic crosses; the intrinsic phenotypes caused by the modifier mutations themselves may thus be assessed in isolation.

Also, such newly-identified modifier mutations may be tested for genetic interactions with other genes-of-interest using methods described above. In particular, modifier genes may be placed into so-called complementation groups, using genetic crosses, for subsequent examination of the phenotypes of progeny that contain two or more modifier mutations. Two modifier mutations are said to fall within the same complementation group if nematodes carrying both mutations exhibit essentially the same phenotype as nematodes carrying each mutation alone. Generally, individual complementation groups defined in this way correspond to individual genes. The precise location and sequence of the modifier gene in the genomic DNA is confirmed by: (a) identifying sequence changes specific to the modifier mutations within the gene in question; and (b) in most cases, demonstrating reversion of the phenotype caused by the modifier mutation upon injection of a limited DNA fragment containing the wild-type form of the modifier gene.

An alternative mutagenesis-and-screening strategy that is especially useful for the rapid identification of modifier genes has also been described (*see* Anderson, 1995,

Mutagenesis, Methods in Cell Biology 4:31-58) which is based on the use of transposable elements as mutagens. Because the mutated modifier gene becomes tagged with sequences derived from the transposable element, such as Tc1 as described above, this strategy allows for easy identification of the modifier gene through PCR amplification of sequences adjacent to the insertion site of the transposon. Mutagenesis may be carried out by introduction of a mutator locus, termed *mut-2*, which promotes mobility of transposons. In this case, the mutator locus is introduced into strains carrying mutations in insulin-like genes, and the progeny examined for suppression or enhancement of the original phenotype, as described above.

Once nematode modifier genes that participate in the same pathway as insulin-like genes have been identified using genetic screens, homologous genes in other species-of-interest can be isolated using procedures based on cross-hybridization with *C. elegans* modifier gene DNA probes, PCR-based strategies with primer sequences derived from those of *C. elegans* modifier genes, and/or computer searches of sequence databases. For therapeutic applications related to the function of insulin superfamily hormones, human and rodent homologs of the nematode modifier genes are of particular interest. For pesticide applications, homologs of nematode modifier genes in agriculturally-important pest species, beneficial insects, and other invertebrate model organisms are of particular interest and include the following: *D. melanogaster, Anopheles, Heliothis virescens, Plodia interpunctella, Spodoptera frugiperda, Pectinophora gosypiella, Plutella xylostella, Tribolium castaneum, Diabrotica spp., Leptinotarsa decemlineata, Anthonomus grandis, Bemisia tabaci, Myzus persicae, Blattella germanica, Apis mellifera, Ctenocephalites felis, Amblyoma americanum, Meloidogyne spp., Heterodera glycinii, etc.*

WHAT IS CLAIMED IS:

1. A method of analyzing an effect of expression or mis-expression of a *C. elegans* insulin-like gene comprising observing a first nematode genetically engineered to express or mis-express a *C. elegans* insulin-like protein of any one of groups I, II or IV, or a derivative or fragment thereof that displays one or more functional activities of the *C. elegans* insulin-like protein.

2. The method of Claim 1, wherein the protein, derivative or fragment comprises an amino acid sequence selected from the group consisting of SEQ ID NOs.1-15, 18, 158-161 and 198-206.

3. The method of Claim 1, wherein the protein, derivative or fragment comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:1, 6, 8, 9, 11, 12, 15, 18, 158-161 and 198-206.

4 The method of Claim 1, wherein the protein, derivative or fragment is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NOs:19-33, 36, 162-165 and 207-215.

5. The method of Claim 1, wherein the protein, derivative or fragment is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NOs:19, 24, 26, 27, 29, 30, 33, 36, 162-165 and 207-215.

6. The method of any of Claims 1-5, wherein the effect is observed in an assay selected from the group consisting of a dauer formation assay, a developmental assay, an energy metabolism assay, a growth rate assay and a reproductive capacity assay.

7. The method of any of Claims 1-5, wherein the *C. elegans* insulin-like protein, derivative or fragment is encoded by a mutated or abnormally expressed gene and the effect observed is the phenotype associated with the mutation or abnormal expression.

8. The method of any of Claims 1-5, wherein the gene encoding the *C. elegans* insulin-like protein, derivative or fragment is caused to be mutated or abnormally expressed.

9. The method of Claim 8, wherein the gene is mutated or abnormally expressed using a technique selected from the group consisting of EMS chemical deletion mutagenesis, transposon insertion mutagenesis and double-stranded RNA interference.

10. The method of Claim 9, further comprising observing a second nematode having the same mutation or abnormal expression in the gene encoding the *C. elegans* insulin-like protein as the first nematode observed, wherein the second nematode additionally comprises a second mutation in a gene-of-interest, and wherein the effect observed is a difference, if any, between the phenotype of the first nematode and the second nematode, wherein a difference in phenotype identifies the gene-of-interest as capable of modifying the function of the gene encoding the *C. elegans* insulin-like protein.

11. The method of Claim 10, wherein the phenotype observed is selected from the group consisting of an altered body shape phenotype, an altered body size phenotype, an altered chemotaxis phenotype, an altered brood size phenotype, an altered egg-laying phenotype, an altered life span phenotype, an altered lipid accumulation phenotype, an altered locomotion phenotype, an altered organ morphogenesis phenotype, an altered thermotaxis phenotype, a dauer constitutive phenotype, a dauer defective phenotype, a lethal phenotype and a sterile phenotype.

12. The method of Claim 11, wherein the phenotype observed is altered organ morphogenesis, and wherein the organ is selected from the group consisting of vulva, nervous system, gut and musculature.

13. The method of Claim 12, wherein the phenotype observed is altered body size, and wherein the nematode is assayed for activity of a gene affecting body size selected from the group consisting of *daf-4*, *sma-2* and *sma-3*.

14. The method of Claim 10, wherein the gene-of-interest is a homolog of an insulin signaling pathway gene from vertebrates.

15. The method of Claim 10, wherein the gene-of-interest is selected from the group consisting of *daf2*, *daf-16* and *age-1*.

16. The method of Claim 15, wherein the gene-of-interest is *daf-2* and the phenotype observed is selected from the group consisting of dauer formation and life span.

17. A purified *C. elegans* insulin-like protein comprising or consisting of an amino acid sequence of any one of SEQ ID NOs:1, 6, 8, 9, 11, 12, 15, 18, 158-161 or 198-206.

18. A purified derivative or fragment of the protein of Claim 17 consisting of at least 10 contiguous amino acids of the *C. elegans* insulin-like protein.

19. The derivative or fragment of Claim 18 which displays one or more functional activities of the *C. elegans* insulin-like protein.

20. The derivative or fragment of Claim 18 which is capable of immunospecific binding to an antibody raised against a *C. elegans* insulin-like protein.

21. A purified molecule comprising the derivative or fragment of any one of Claims 18-20.

22. A chimeric protein comprising a fragment of the *C. elegans* insulin-like protein of Claim 17 consisting of at least 10 contiguous amino acids of the *C. elegans* insulin-like protein fused by a covalent bond to an amino acid sequence of a second protein, which second protein is not a *C. elegans* insulin-like protein.

23. A purified antibody or an antigen-binding fragment or derivative thereof capable of immunospecific binding to the protein, derivative or fragment of any one of Claims 17-20 and not to an insulin-like protein of another species.

24. A composition comprising the protein, derivative or fragment of any one of Claims 17-20 and a pharmaceutically acceptable carrier.

25. The protein of Claim 17, wherein the protein further comprises a domain depicted in any of FIGs. 4-34, wherein the domain is selected from the group consisting of a signal peptide, a pro peptide, an A domain, a B domain, and a C domain.

26. The protein of Claim 17, wherein the protein further comprises a B peptide domain linked by one or more disulfide bonds to an A peptide domain.

27. The protein of Claim 26, wherein said B and A peptide domains have not been proteolytically cleaved into separate chains.

28. A mature *C. elegans* insulin-like protein which is the result of expressing a nucleic acid encoding the protein of Claim 17.

29. The protein of Claim 17, wherein the *C. elegans* insulin-like protein is a Class IV protein.

30. An isolated nucleic acid or a complement thereof which comprises a heterologous nucleotide sequence of less than 15,000 nucleotides that encodes at least 10 contiguous amino acids of a *C. elegans* insulin-like protein of Claim 17, provided that the isolated nucleic acid is not a cosmid.

31. The isolated nucleic acid of Claim 30, which comprises a nucleotide sequence of any one of SEQ ID NOs:19, 24, 26, 27, 29, 30, 33, 36, 162-165 and 207-215, or which encodes a *C. elegans* insulin-like protein comprising any one of SEQ ID NOs:1, 6, 8, 9, 11, 12, 15, 18, 158-161 and 198-206.

32 The isolated nucleic acid of Claim 30, which encodes one or more domains as annotated and defined by an amino acid sequence depicted in any of FIGs. 4-34, wherein the domain is selected from the group consisting of a signal peptide, a pro peptide, an A domain, a B domain, and a C domain.

33. The isolated nucleic acid of Claim 30, further comprising a nucleotide sequence encoding a functional derivative of at least a portion of an amino acid sequence selected from the group consisting of any one of SEQ ID NOs:1-15, 18, 158-161 and 198-206.

34. A non-human animal comprising a transgene which encodes a *C. elegans* insulin-like protein, derivative or fragment of Claim 18.

35. The non-human animal of Claim 34 which is a *C. elegans* animal and further comprises at least one deleted or inactivated *C. elegans* insulin-like gene encoding an amino acid sequence selected from the group consisting of SEQ ID NOs:1-18, 158-161 and 198-206.

36. The method of any of Claims 1-5, wherein the expression of the *C. elegans* insulin-like protein is driven by a heterologous promoter.

37. The method of Claim 36, wherein the heterologous promoter is selected from the group consisting of an *hsp 16-2* promoter, an *hsp 16-41* promoter, a *myo-2* promoter, an *hlh-1* promoter and a *mec-3* promoter.

38. The method of Claim 37, additionally comprising contacting the nematode with one or more molecules and determining whether the one or more molecules alters the expression of the *C. elegans* insulin-like protein.

39. An isolated nucleic acid or a complement thereof which comprises a heterologous nucleotide sequence of less than 500 nucleotides that encodes at least 10 contiguous amino acids of a *C. elegans* insulin-like protein of Claim 17.

40. A purified *C. elegans* insulin-like protein of any one of groups I, II or IV, or a derivative or fragment thereof that displays one or more functional activities of the *C. elegans* insulin-like protein, for use in insulin-related research.

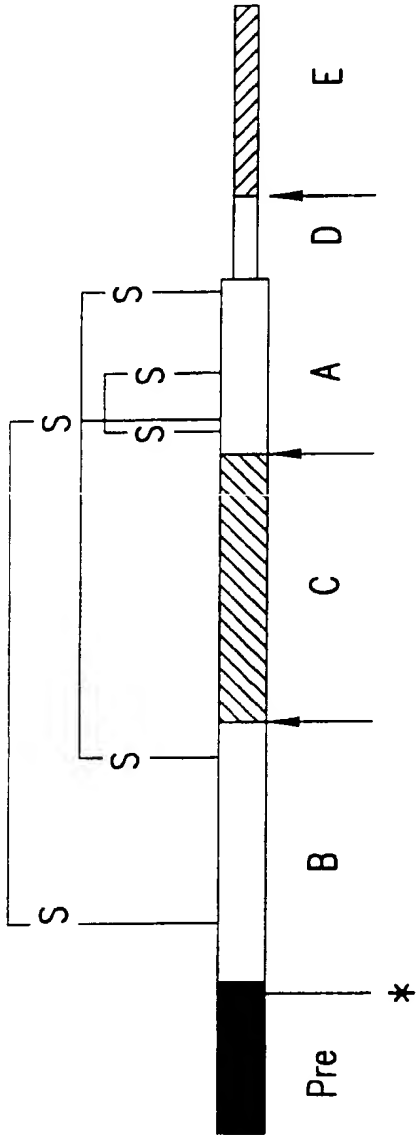


FIG.1

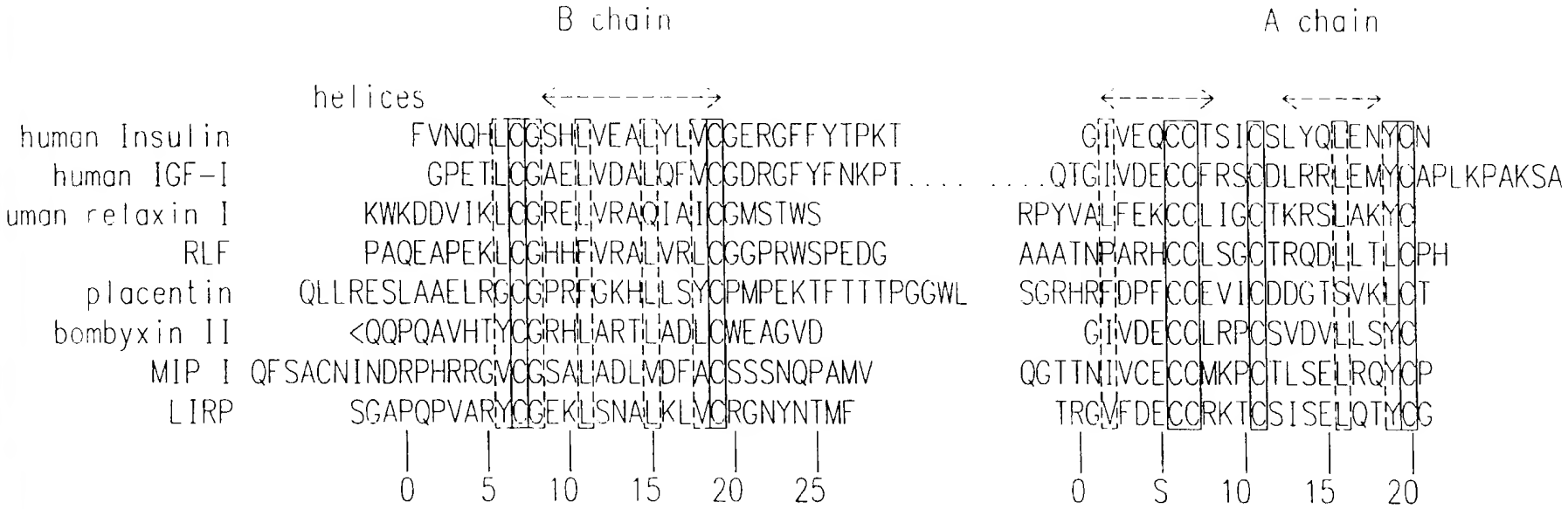

B chain
A chain
helices
human Insulin
human IGF-I
uman relaxin I
RLF
placentin
bombyxin II
MIP I
LIRP
0
5
10
15
20
25
S

FIG.2

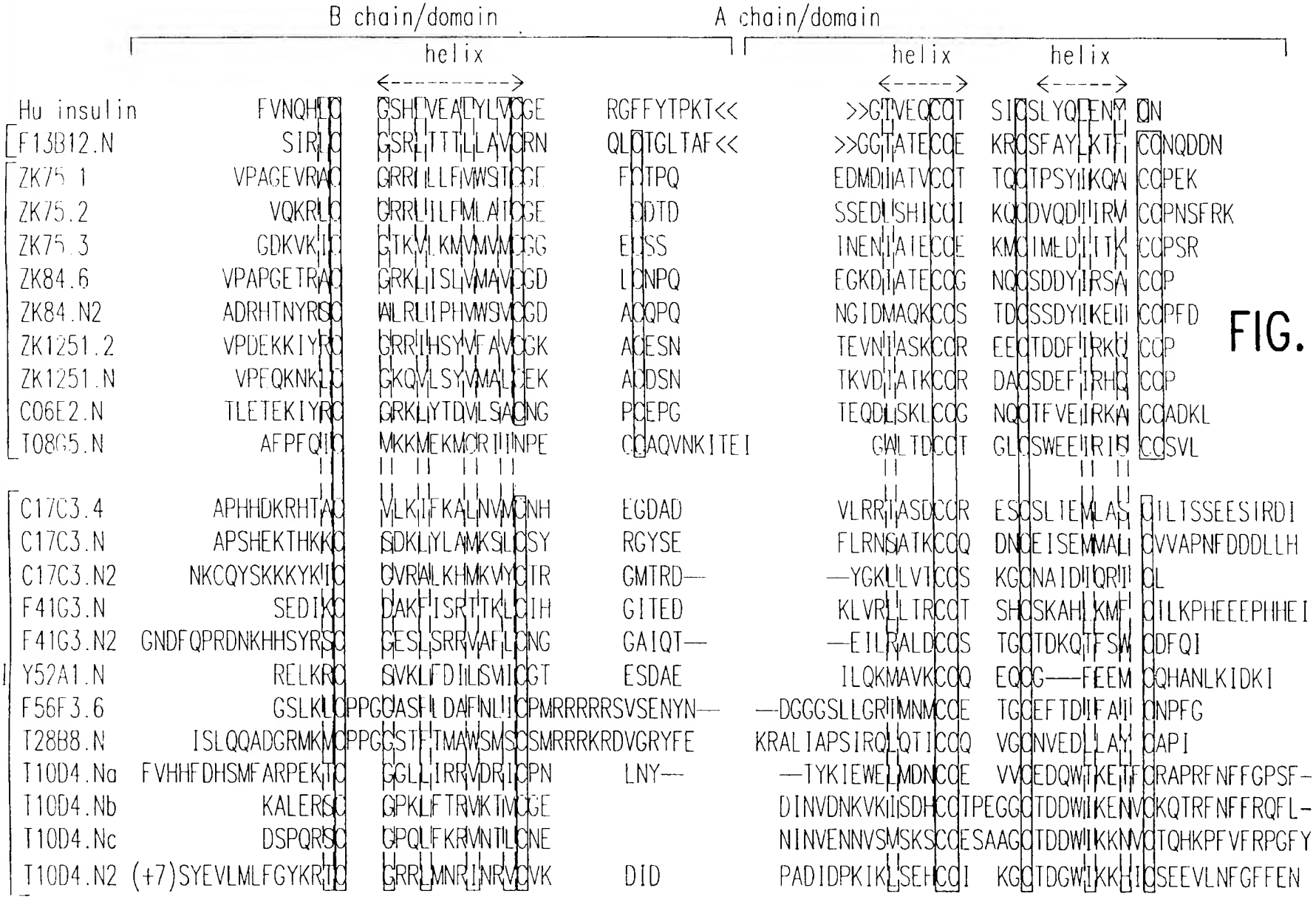

B chain/domain
A chain/domain
helix
helix
helix
Hu insulin
F13B12.N
ZK75.1
ZK75.2
ZK75.3
ZK84.6
ZK84.N2
ZK1251.2
ZK1251.N
C06E2.N
T08G5.N
C17C3.4
C17C3.N
C17C3.N2
F41G3.N
F41G3.N2
Y52A1.N
F56F3.6
T28B8.N
T10D4.Na
T10D4.Nb
T10D4.Nc
T10D4.N2

FIG. 3A

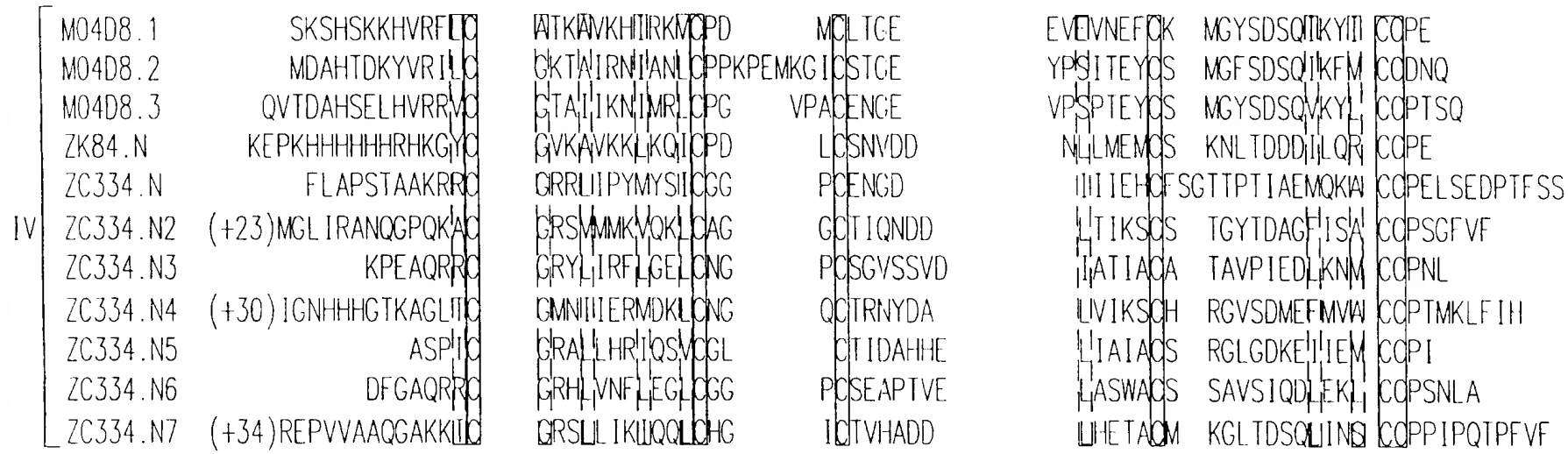

IV
M04D8.1 SKSHSKKHVRFLC ATKAVKHIIRKMCPD MCLTGE EVEVNEFCK MGYSDSQTKYIICQPE
M04D8.2 MDAHTDKYVRILC GKTAIRNIIANLCPPKPEMKGICSTGE YPSITEYCS MGFSDSQIKFMCQDNQ
M04D8.3 QVTDAHSELHVRRVC GTAIIKNIMRLCPG VPACENGE VPSPTEYCS MGYSDSQVKYLCQPTSQ
ZK84.N KEPKHHHHHHHRHKGYC GVKAVKKLKQICPD LCSNVDD NLLMEMCS KNLTDDDILQRCQPE
ZC334.N FLAPSTAAKRRC GRRLIIPYMYSICCG PCENGD IIIIEHCFSGTTPTIAEMQKAVCQPELSEDPTFSS
ZC334.N2 (+23)MGLIRANQGPQKAC GRSVMMKVQKLCAG GCTIQNDD LTIKSCS TGYTDAGFISACQPSGFVF
ZC334.N3 KPEAQRRC GRYLIRFLGELCNG PCSGVSSVD IATIACA TAVPIEDLKNMCQPNL
ZC334.N4 (+30)IGNHHHGTKAGLTC GMNIIIERMDKLCNG QCTRNYDA LIVIKSCH RGVSDMEFMVWCQPTMKLFIH
ZC334.N5 ASPIC GRALLHRIQSVCGL CTIDAHHE LIAIACS RGLGDKEIIEMCQPI
ZC334.N6 DFGAQRRC GRHLVNFLEGLCGG PCSEAPTVE LASWACS SAVSIQDLEKLCQPSNLA
ZC334.N7 (+34)REPVVAAQGAKKIC GRSLLIKLIQQLCHG ICTVHADD LHETACM KGLTDSQLINSCQPPIPQTPFVF

FIG.3B

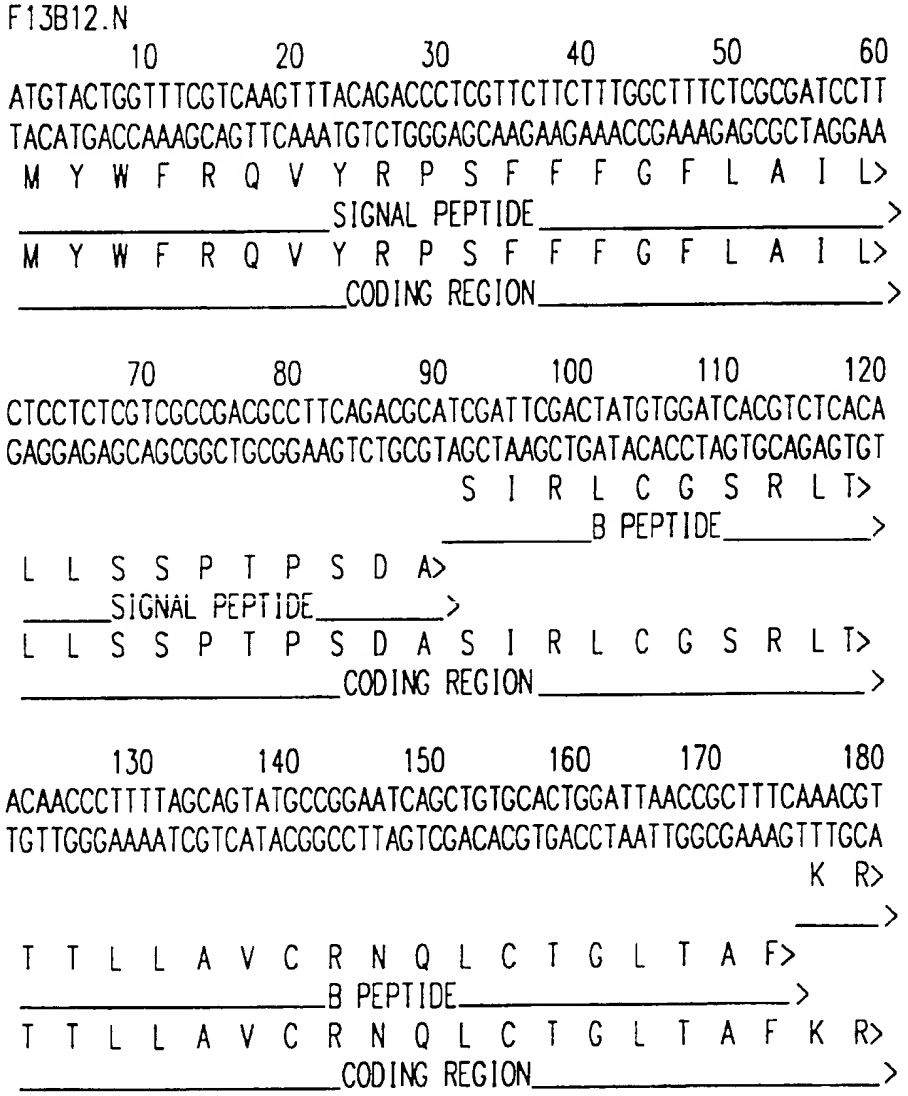

F13B12.N
10 20 30 40 50 60
ATGTACTGGTTTCGTCAAGTTTACAGACCCTCGTTCTTCTTTGGCTTTCTCGCGATCCTT
TACATGACCAAAGCAGTTCAAATGTCTGGGAGCAAGAAGAAACCGAAAGAGCGCTAGGAA
M Y W F R Q V Y R P S F F F G F L A I L>
SIGNAL PEPTIDE
M Y W F R Q V Y R P S F F F G F L A I L>
CODING REGION
70 80 90 100 110 120
CTCCTCTCGTCGCCGACGCCTTCAGACGCATCGATTCGACTATGTGGATCACGTCTCACA
GAGGAGAGCAGCGGCTGCGGAAGTCTGCGTAGCTAAGCTGATACACCTAGTGCAGAGTGT
S I R L C G S R L T>
B PEPTIDE
L L S S P T P S D A>
SIGNAL PEPTIDE
L L S S P T P S D A S I R L C G S R L T>
CODING REGION
130 140 150 160 170 180
ACAACCCTTTTAGCAGTATGCCGGAATCAGCTGTGCACTGGATTAACCGCTTTCAAACGT
TGTTGGGAAAATCGTCATACGGCCTTAGTCGACACGTGACCTAATTGGCGAAAGTTTGCA
K R>
T T L L A V C R N Q L C T G L T A F>
B PEPTIDE
T T L L A V C R N Q L C T G L T A F K R>
CODING REGION

FIG. 4A

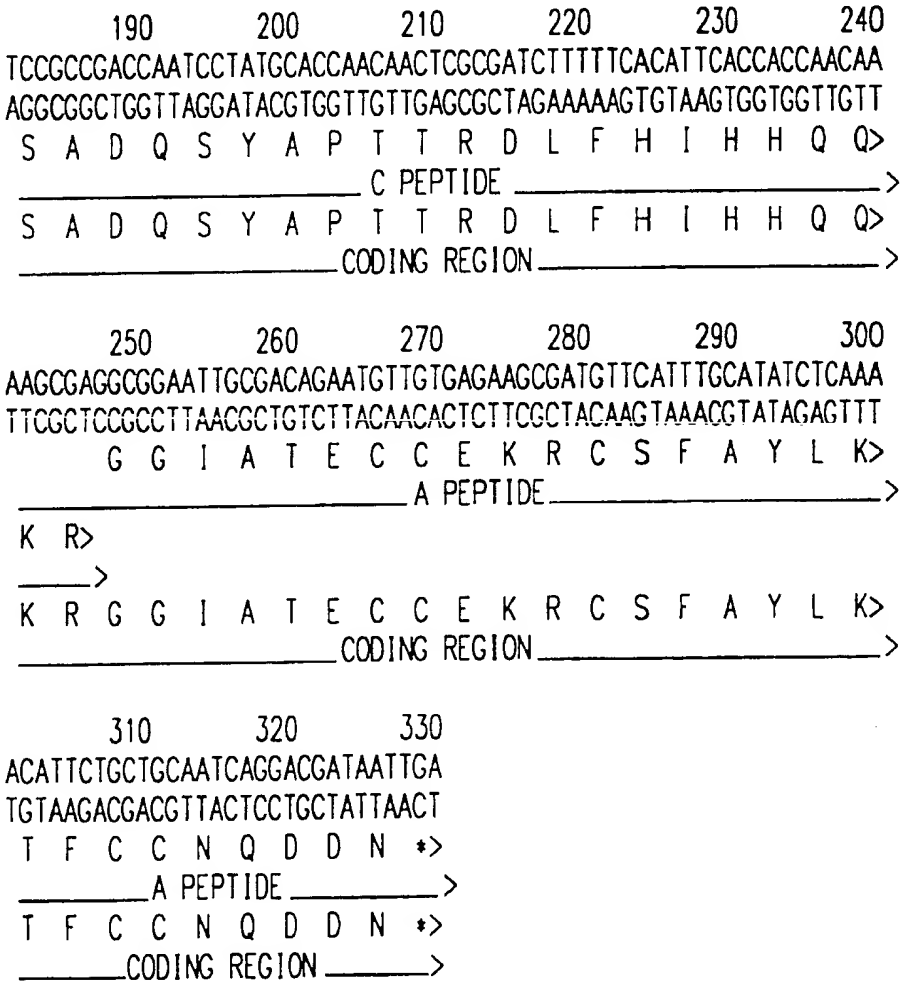

190 200 210 220 230 240
TCCGCCGACCAATCCTATGCACCAACAACTCGCGATCTTTTTCACATTCACCACCAACAA
AGGCGGCTGGTTAGGATACGTGGTTGTTGAGCGCTAGAAAAAGTGTAAGTGGTGGTTGTT
S A D Q S Y A P T T R D L F H I H H Q Q>
C PEPTIDE
S A D Q S Y A P T T R D L F H I H H Q Q>
CODING REGION
250 260 270 280 290 300
AAGCGAGGCGGAATTGCGACAGAATGTTGTGAGAAGCGATGTTCATTTGCATATCTCAAA
TTCGCTCCGCCTTAACGCTGTCTTACAACACTCTTCGCTACAAGTAAACGTATAGAGTTT
G G I A T E C C E K R C S F A Y L K>
A PEPTIDE
K R>
K R G G I A T E C C E K R C S F A Y L K>
CODING REGION
310 320 330
ACATTCTGCTGCAATCAGGACGATAATTGA
TGTAAGACGACGTTACTCCTGCTATTAACT
T F C C N Q D D N *>
A PEPTIDE
T F C C N Q D D N *>
CODING REGION

FIG. 4B

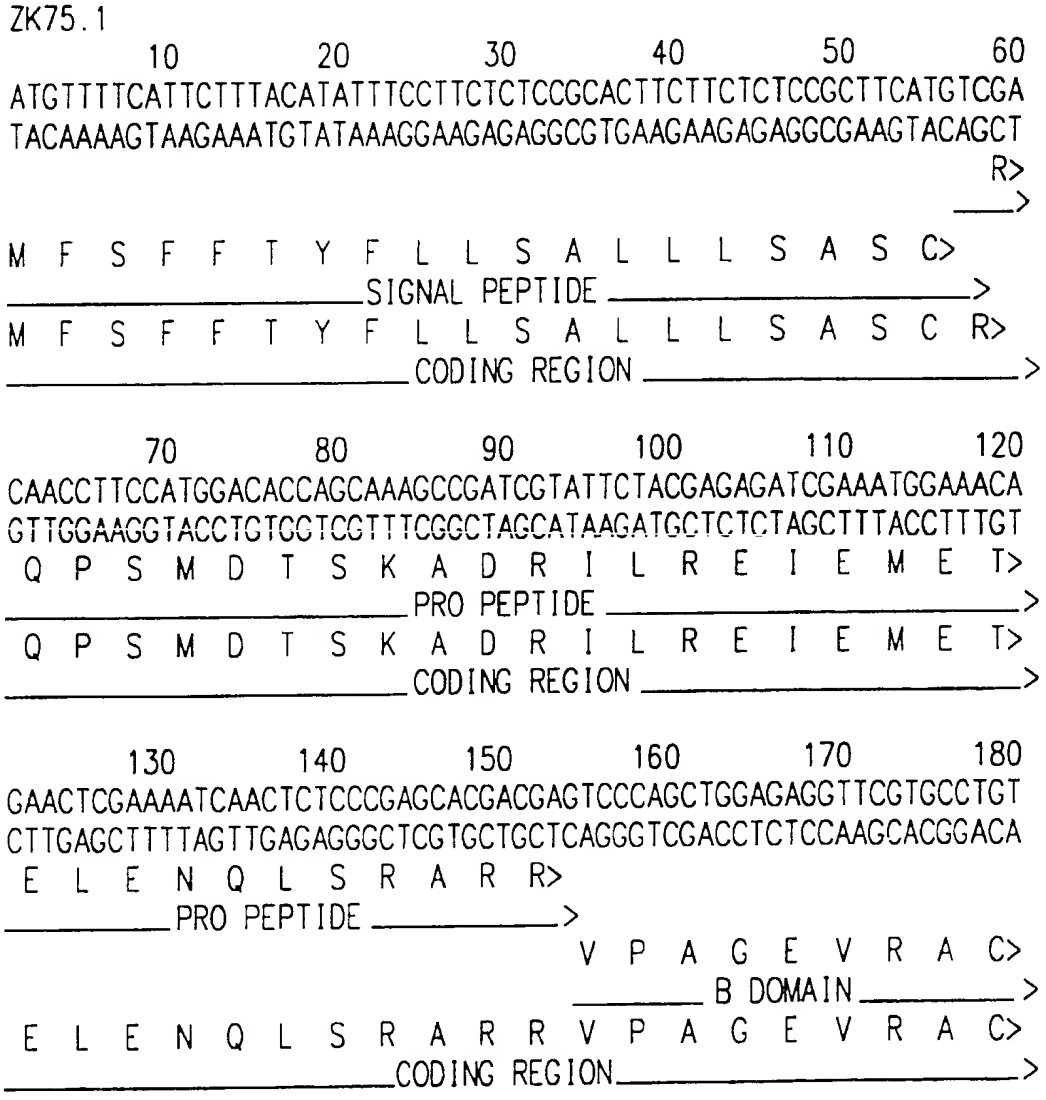

ZK75.1
10 20 30 40 50 60
ATGTTTTCATTCTTTACATATTTCCTTCTCTCCGCACTTCTTCTCTCCGCTTCATGTCGA
TACAAAAGTAAGAAATGTATAAAGGAAGAGAGGCGTGAAGAAGAGAGGCGAAGTACAGCT
R>
M F S F F T Y F L L S A L L L S A S C>
SIGNAL PEPTIDE
M F S F F T Y F L L S A L L L S A S C R>
CODING REGION
70 80 90 100 110 120
CAACCTTCCATGGACACCAGCAAAGCCGATCGTATTCTACGAGAGATCGAAATGGAAACA
GTTGGAAGGTACCTGTGGTCGTTTCGGCTAGCATAAGATGCTCTCTAGCTTTACCTTTGT
Q P S M D T S K A D R I L R E I E M E T>
PRO PEPTIDE
Q P S M D T S K A D R I L R E I E M E T>
CODING REGION
130 140 150 160 170 180
GAACTCGAAAATCAACTCTCCCGAGCACGACGAGTCCCAGCTGGAGAGGTTCGTGCCTGT
CTTGAGCTTTTAGTTGAGAGGGCTCGTGCTGCTCAGGGTCGACCTCTCCAAGCACGGACA
E L E N Q L S R A R R>
PRO PEPTIDE
V P A G E V R A C>
B DOMAIN
E L E N Q L S R A R R V P A G E V R A C>
CODING REGION

FIG. 5A

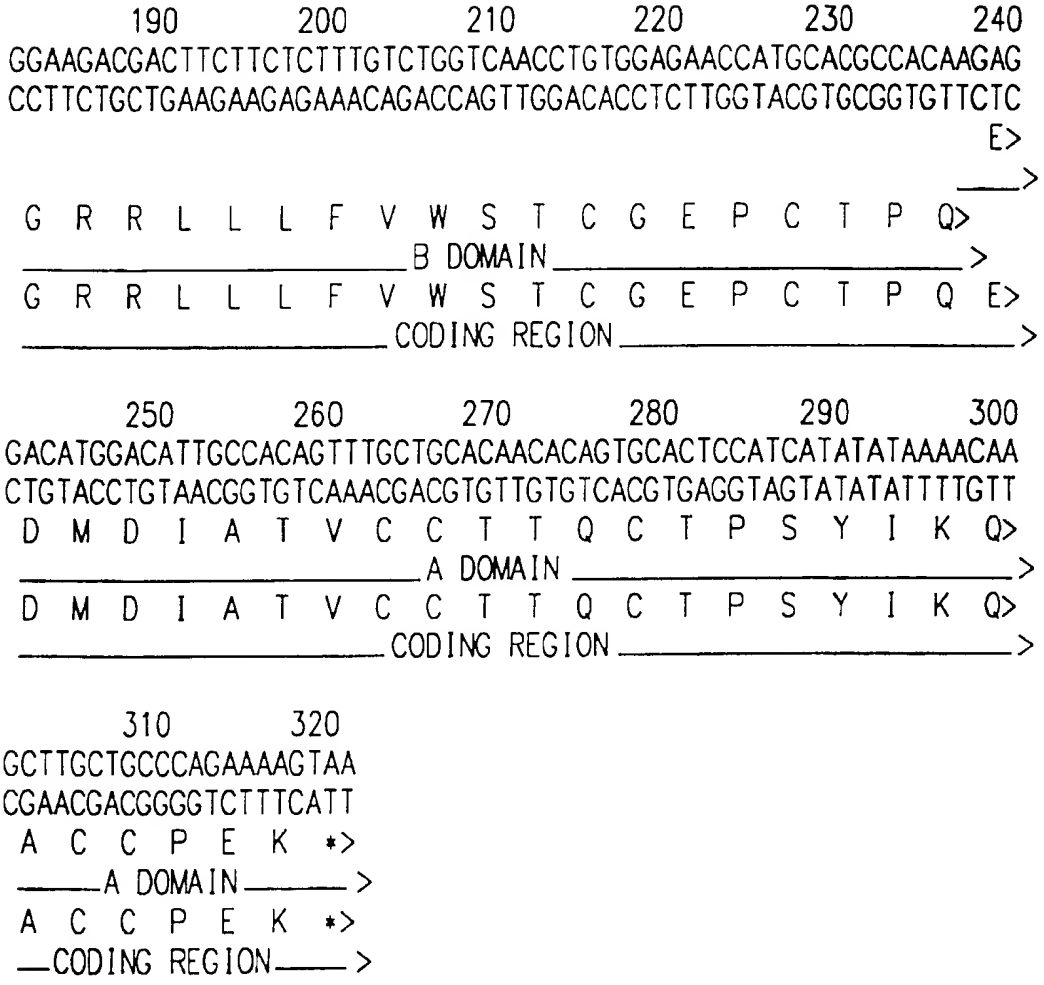

190 200 210 220 230 240
GGAAGACGACTTCTTCTCTTTGTCTGGTCAACCTGTGGAGAACCATGCACGCCACAAGAG
CCTTCTGCTGAAGAAGAGAAACAGACCAGTTGGACACCTCTTGGTACGTGCGGTGTTCTC
E>
G R R L L L F V W S T C G E P C T P Q>
B DOMAIN
G R R L L L F V W S T C G E P C T P Q E>
CODING REGION
250 260 270 280 290 300
GACATGGACATTGCCACAGTTTGCTGCACAACACAGTGCACTCCATCATATATAAAACAA
CTGTACCTGTAACGGTGTCAAACGACGTGTTGTGTCACGTGAGGTAGTATATATTTTGTT
D M D I A T V C C T T Q C T P S Y I K Q>
A DOMAIN
D M D I A T V C C T T Q C T P S Y I K Q>
CODING REGION
310 320
GCTTGCTGCCCAGAAAAGTAA
CGAACGACGGGGTCTTTCATT
A C C P E K *>
A DOMAIN
A C C P E K *>
CODING REGION

FIG. 5B

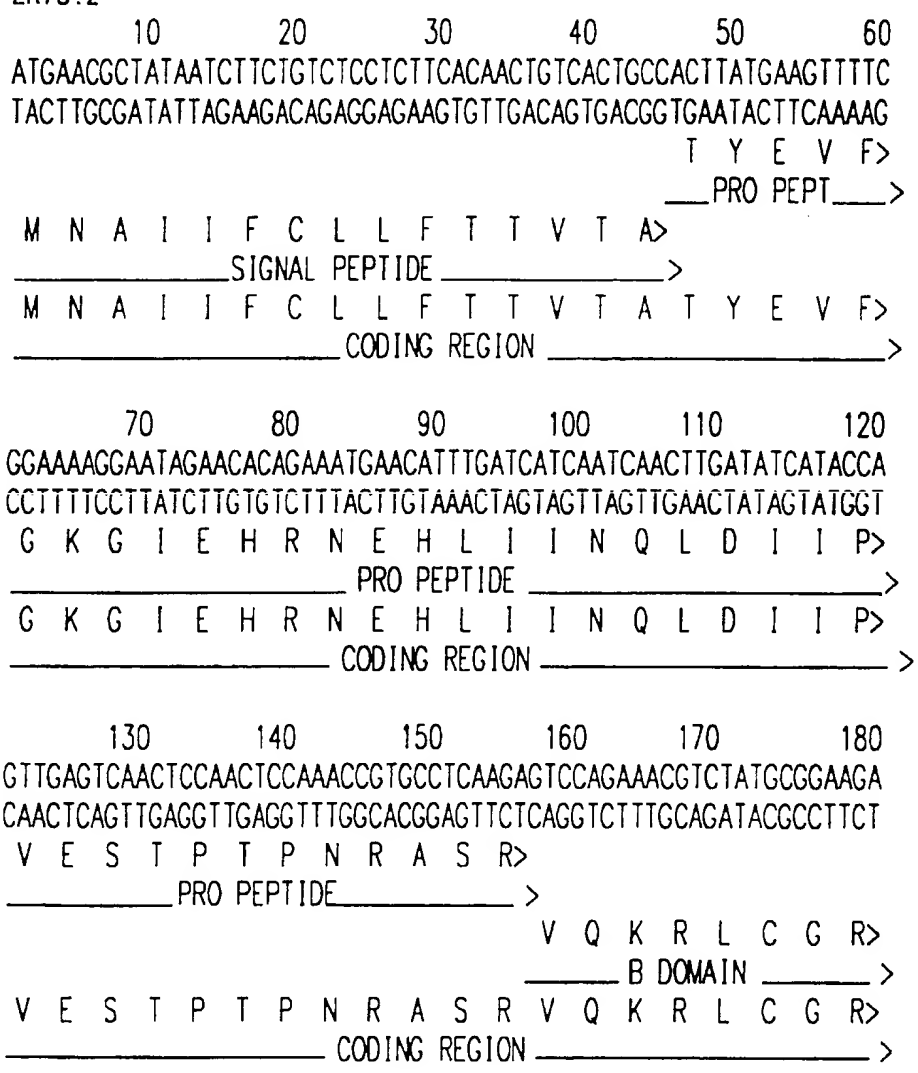

ZK75.2
10 20 30 40 50 60
ATGAACGCTATAATCTTCTGTCTCCTCTTCACAACTGTCACTGCCACTTATGAAGTTTTC
TACTTGCGATATTAGAAGACAGAGGAGAAGTGTTGACAGTGACGGTGAATACTTCAAAAG
T Y E V F>
PRO PEPT
M N A I I F C L L F T T V T A>
SIGNAL PEPTIDE
M N A I I F C L L F T T V T A T Y E V F>
CODING REGION
70 80 90 100 110 120
GGAAAAGGAATAGAACACAGAAATGAACATTTGATCATCAATCAACTTGATATCATACCA
CCTTTTCCTTATCTTGTGTCTTTACTTGTAAACTAGTAGTTAGTTGAACTATAGTATGGT
G K G I E H R N E H L I I N Q L D I I P>
PRO PEPTIDE
G K G I E H R N E H L I I N Q L D I I P>
CODING REGION
130 140 150 160 170 180
GTTGAGTCAACTCCAACTCCAAACCGTGCCTCAAGAGTCCAGAAACGTCTATGCGGAAGA
CAACTCAGTTGAGGTTGAGGTTTGGCACGGAGTTCTCAGGTCTTTGCAGATACGCCTTCT
V E S T P T P N R A S R>
PRO PEPTIDE
V Q K R L C G R>
B DOMAIN
V E S T P T P N R A S R V Q K R L C G R>
CODING REGION

FIG. 6A

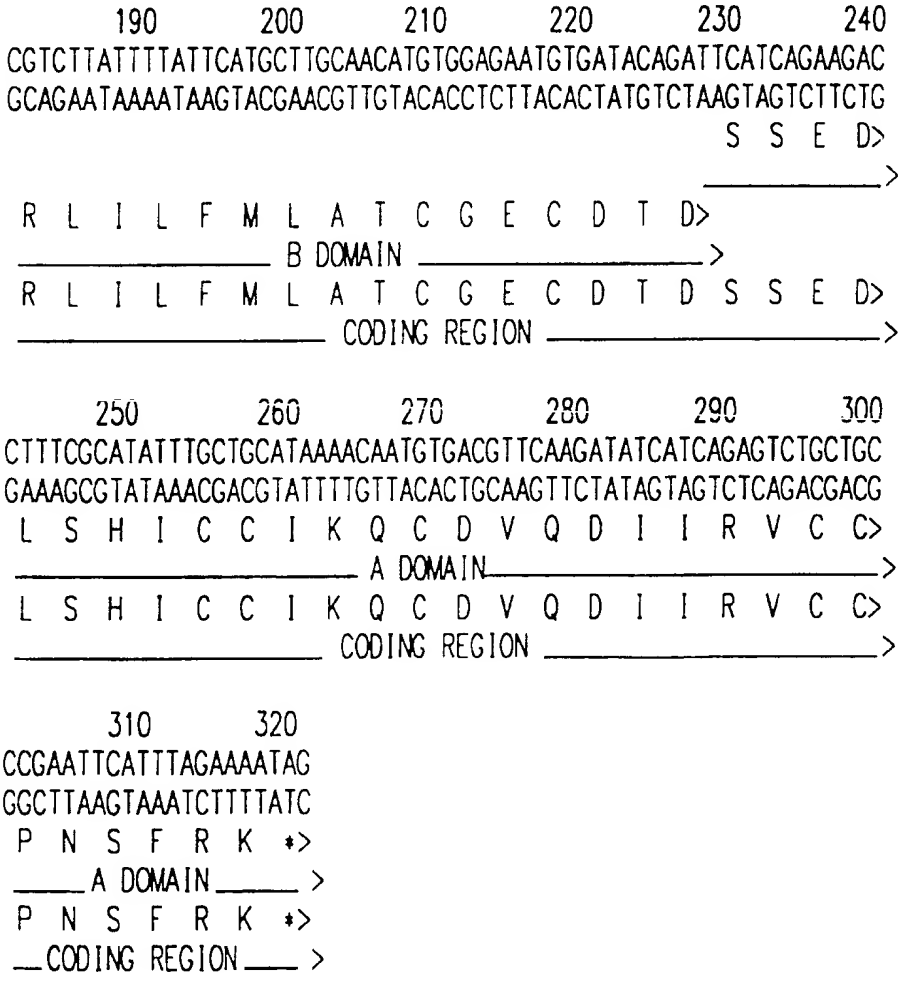

190 200 210 220 230 240
CGTCTTATTTTATTCATGCTTGCAACATGTGGAGAATGTGATACAGATTCATCAGAAGAC
GCAGAATAAAATAAGTACGAACGTTGTACACCTCTTACACTATGTCTAAGTAGTCTTCTG
S S E D>
R L I L F M L A T C G E C D T D>
B DOMAIN
R L I L F M L A T C G E C D T D S S E D>
CODING REGION
250 260 270 280 290 300
CTTTCGCATATTTGCTGCATAAAACAATGTGACGTTCAAGATATCATCAGAGTCTGCTGC
GAAAGCGTATAAACGACGTATTTTGTTACACTGCAAGTTCTATAGTAGTCTCAGACGACG
L S H I C C I K Q C D V Q D I I R V C C>
A DOMAIN
L S H I C C I K Q C D V Q D I I R V C C>
CODING REGION
310 320
CCGAATTCATTTAGAAAATAG
GGCTTAAGTAAATCTTTTATC
P N S F R K *>
A DOMAIN
P N S F R K *>
CODING REGION

FIG. 6B

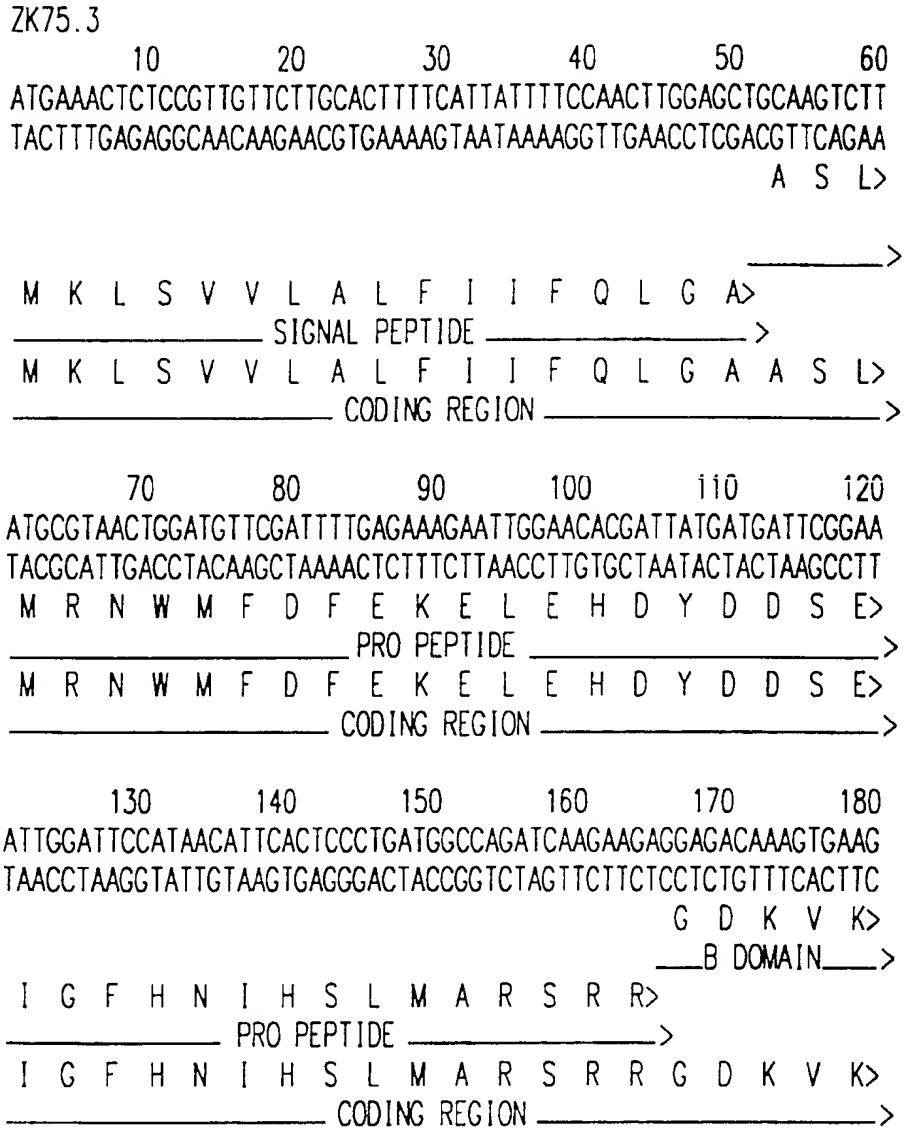

ZK75.3
10 20 30 40 50 60
ATGAAACTCTCCGTTGTTCTTGCACTTTTCATTATTTTCCAACTTGGAGCTGCAAGTCTT
TACTTTGAGAGGCAACAAGAACGTGAAAAGTAATAAAAGGTTGAACCTCGACGTTCAGAA
A S L>
M K L S V V L A L F I I F Q L G A>
SIGNAL PEPTIDE
M K L S V V L A L F I I F Q L G A A S L>
CODING REGION
70 80 90 100 110 120
ATGCGTAACTGGATGTTCGATTTTGAGAAAGAATTGGAACACGATTATGATGATTCGGAA
TACGCATTGACCTACAAGCTAAAACTCTTTCTTAACCTTGTGCTAATACTACTAAGCCTT
M R N W M F D F E K E L E H D Y D D S E>
PRO PEPTIDE
M R N W M F D F E K E L E H D Y D D S E>
CODING REGION
130 140 150 160 170 180
ATTGGATTCCATAACATTCACTCCCTGATGGCCAGATCAAGAAGAGGAGACAAAGTGAAG
TAACCTAAGGTATTGTAAGTGAGGGACTACCGGTCTAGTTCTTCTCCTCTGTTTCACTTC
G D K V K>
B DOMAIN
I G F H N I H S L M A R S R R>
PRO PEPTIDE
I G F H N I H S L M A R S R R G D K V K>
CODING REGION

FIG. 7A

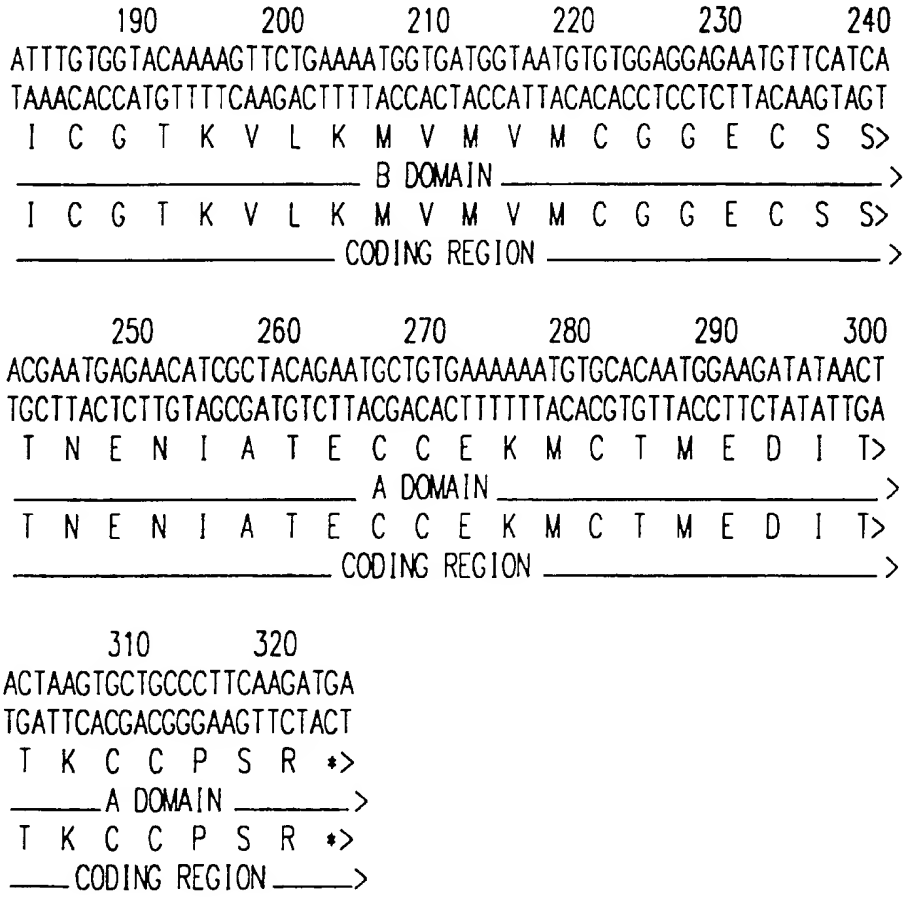

190 200 210 220 230 240
ATTTGTGGTACAAAAGTTCTGAAAATGGTGATGGTAATGTGTGGAGGAGAATGTTCATCA
TAAACACCATGTTTTCAAGACTTTTACCACTACCATTACACACCTCCTCTTACAAGTAGT
I C G T K V L K M V M V M C G G E C S S>
B DOMAIN
I C G T K V L K M V M V M C G G E C S S>
CODING REGION
250 260 270 280 290 300
ACGAATGAGAACATCGCTACAGAATGCTGTGAAAAAATGTGCACAATGGAAGATATAACT
TGCTTACTCTTGTAGCGATGTCTTACGACACTTTTTTACACGTGTTACCTTCTATATTGA
T N E N I A T E C C E K M C T M E D I T>
A DOMAIN
T N E N I A T E C C E K M C T M E D I T>
CODING REGION
310 320
ACTAAGTGCTGCCCTTCAAGATGA
TGATTCACGACGGGAAGTTCTACT
T K C C P S R *>
A DOMAIN
T K C C P S R *>
CODING REGION

FIG. 7B

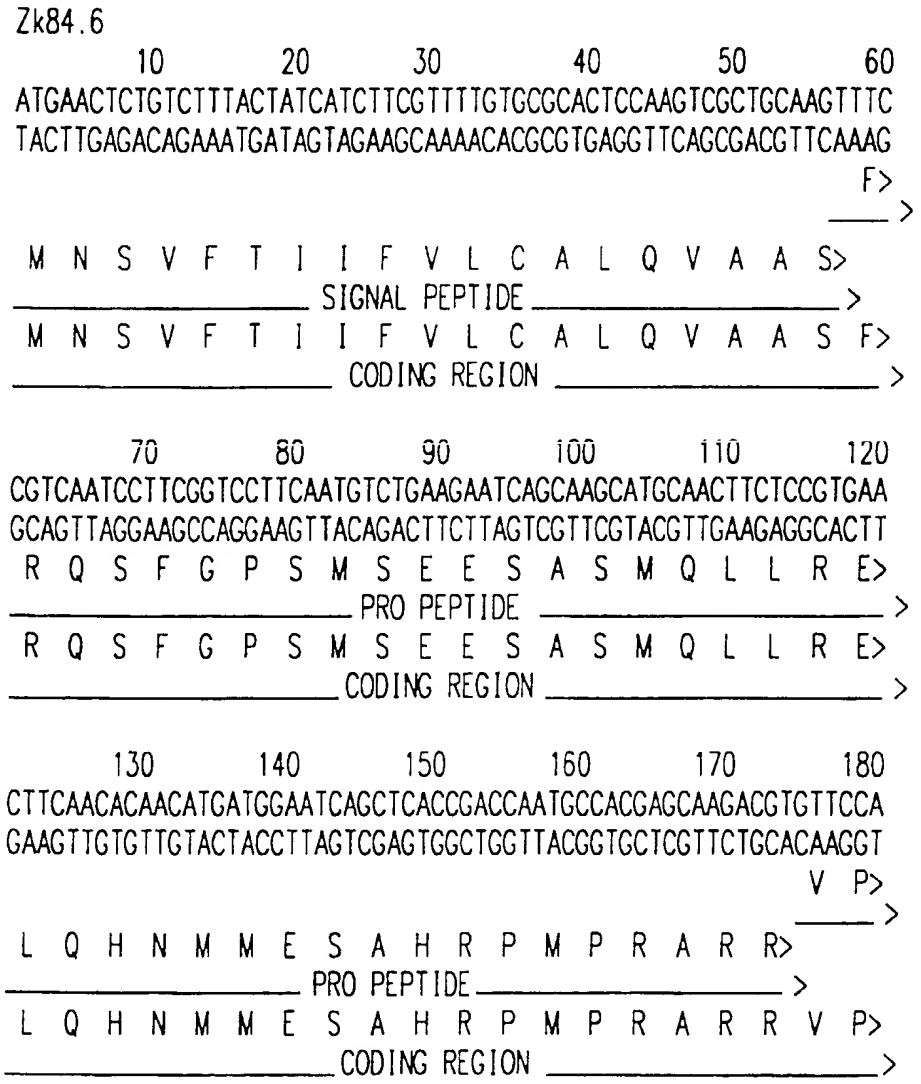

Zk84.6
10 20 30 40 50 60
ATGAACTCTGTCTTTACTATCATCTTCGTTTTGTGCGCACTCCAAGTCGCTGCAAGTTTC
TACTTGAGACAGAAATGATAGTAGAAGCAAAACACGCGTGAGGTTCAGCGACGTTCAAAG
F>
M N S V F T I I F V L C A L Q V A A S>
SIGNAL PEPTIDE
M N S V F T I I F V L C A L Q V A A S F>
CODING REGION
70 80 90 100 110 120
CGTCAATCCTTCGGTCCTTCAATGTCTGAAGAATCAGCAAGCATGCAACTTCTCCGTGAA
GCAGTTAGGAAGCCAGGAAGTTACAGACTTCTTAGTCGTTCGTACGTTGAAGAGGCACTT
R Q S F G P S M S E E S A S M Q L L R E>
PRO PEPTIDE
R Q S F G P S M S E E S A S M Q L L R E>
CODING REGION
130 140 150 160 170 180
CTTCAACACAACATGATGGAATCAGCTCACCGACCAATGCCACGAGCAAGACGTGTTCCA
GAAGTTGTGTTGTACTACCTTAGTCGAGTGGCTGGTTACGGTGCTCGTTCTGCACAAGGT
V P>
L Q H N M M E S A H R P M P R A R R>
PRO PEPTIDE
L Q H N M M E S A H R P M P R A R R V P>
CODING REGION

FIG. 8A

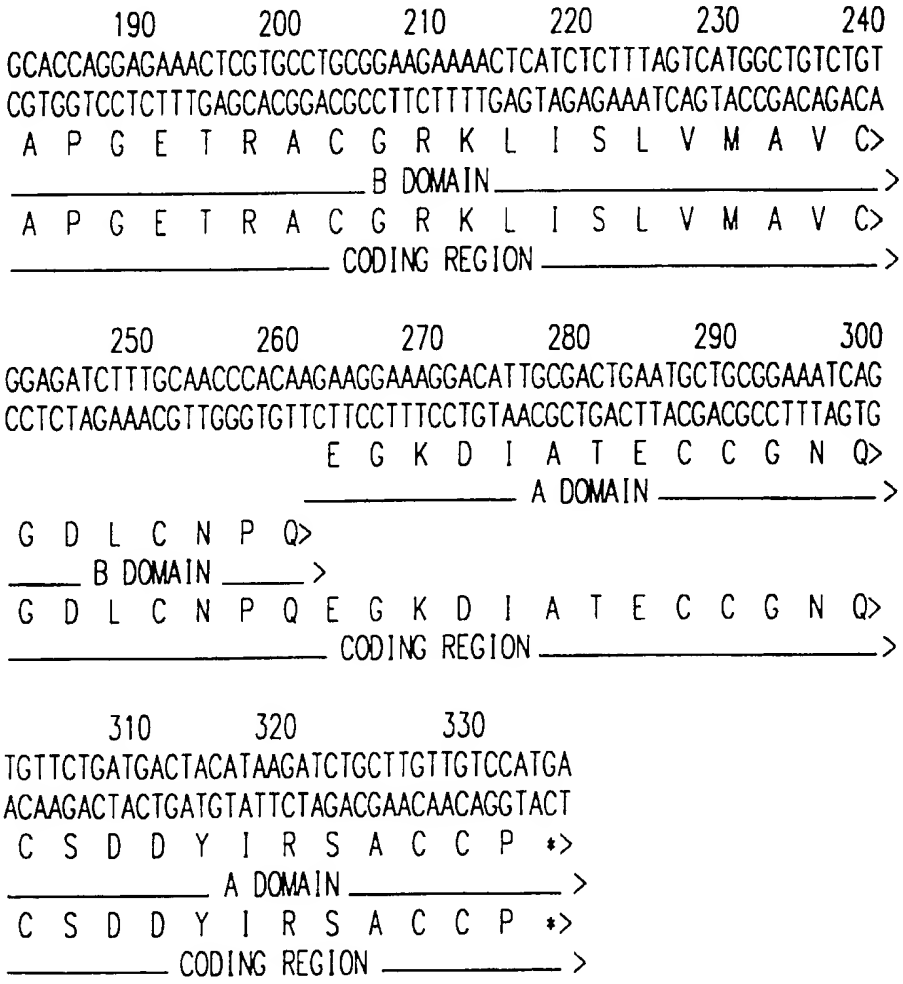

190 200 210 220 230 240
GCACCAGGAGAAACTCGTGCCTGCGGAAGAAAACTCATCTCTTTAGTCATGGCTGTCTGT
CGTGGTCCTCTTTGAGCACGGACGCCTTCTTTTGAGTAGAGAAATCAGTACCGACAGACA
A P G E T R A C G R K L I S L V M A V C>
B DOMAIN
A P G E T R A C G R K L I S L V M A V C>
CODING REGION
250 260 270 280 290 300
GGAGATCTTTGCAACCCACAAGAAGGAAAGGACATTGCGACTGAATGCTGCGGAAATCAG
CCTCTAGAAACGTTGGGTGTTCTTCCTTTCCTGTAACGCTGACTTACGACGCCTTTAGTG
E G K D I A T E C C G N Q>
A DOMAIN
G D L C N P Q>
B DOMAIN
G D L C N P Q E G K D I A T E C C G N Q>
CODING REGION
310 320 330
TGTTCTGATGACTACATAAGATCTGCTTGTTGTCCATGA
ACAAGACTACTGATGTATTCTAGACGAACAACAGGTACT
C S D D Y I R S A C C P *>
A DOMAIN
C S D D Y I R S A C C P *>
CODING REGION

FIG. 8B

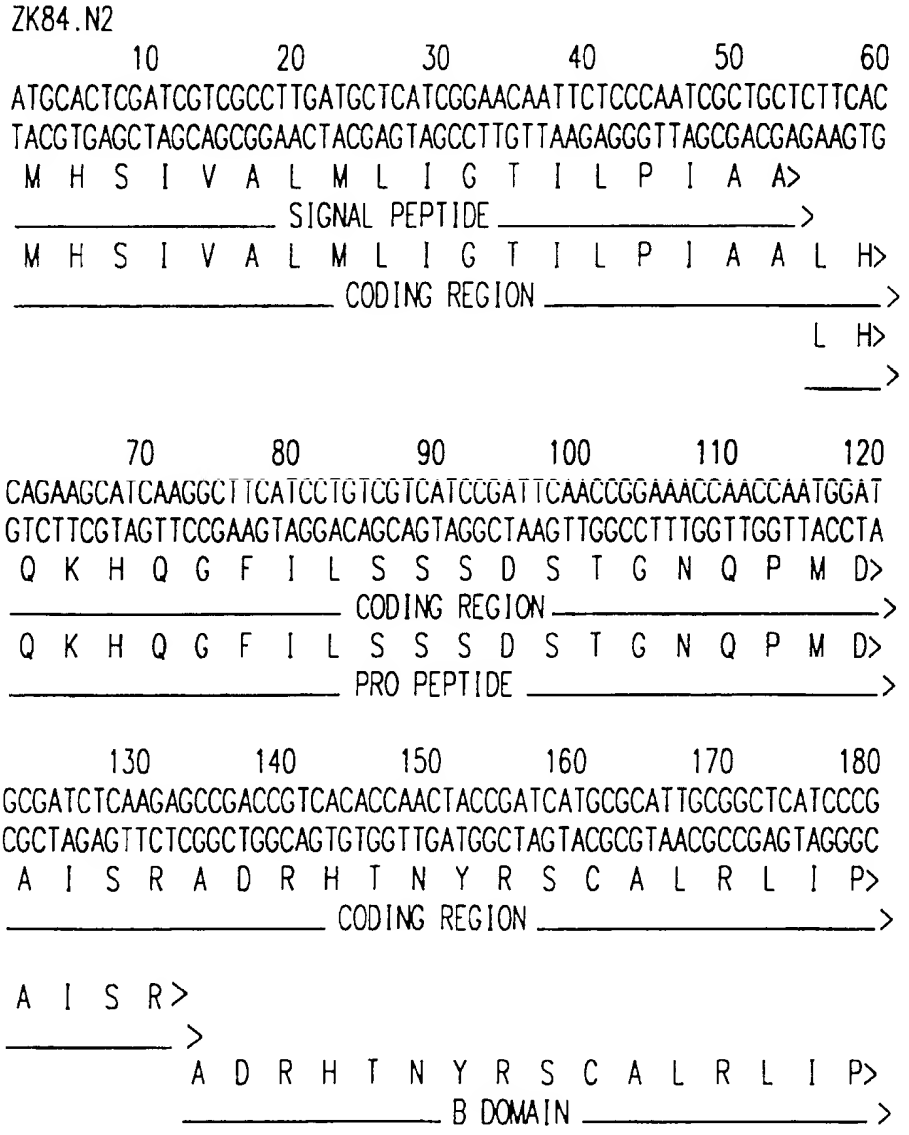

ZK84.N2
10 20 30 40 50 60
ATGCACTCGATCGTCGCCTTGATGCTCATCGGAACAATTCTCCCAATCGCTGCTCTTCAC
TACGTGAGCTAGCAGCGGAACTACGAGTAGCCTTGTTAAGAGGGTTAGCGACGAGAAGTG
M H S I V A L M L I G T I L P I A A>
SIGNAL PEPTIDE
M H S I V A L M L I G T I L P I A A L H>
CODING REGION
L H>
70 80 90 100 110 120
CAGAAGCATCAAGGCTTCATCCTGTCGTCATCCGATTCAACCGGAAACCAACCAATGGAT
GTCTTCGTAGTTCCGAAGTAGGACAGCAGTAGGCTAAGTTGGCCTTTGGTTGGTTACCTA
Q K H Q G F I L S S S D S T G N Q P M D>
CODING REGION
Q K H Q G F I L S S S D S T G N Q P M D>
PRO PEPTIDE
130 140 150 160 170 180
GCGATCTCAAGAGCCGACCGTCACACCAACTACCGATCATGCGCATTGCGGCTCATCCCG
CGCTAGAGTTCTCGGCTGGCAGTGTGGTTGATGGCTAGTACGCGTAACGCCGAGTAGGGC
A I S R A D R H T N Y R S C A L R L I P>
CODING REGION
A I S R>
A D R H T N Y R S C A L R L I P>
B DOMAIN

FIG. 9A

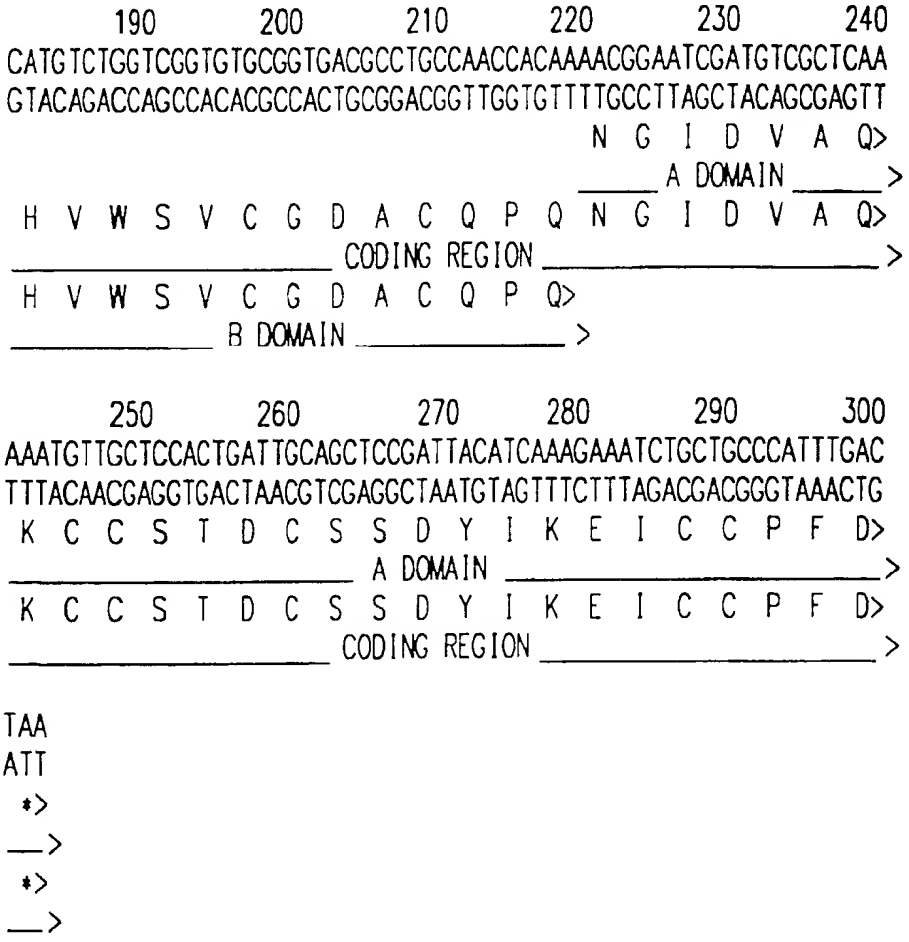

190 200 210 220 230 240
CATGTCTGGTCGGTGTGCGGTGACGCCTGCCAACCACAAAACGGAATCGATGTCGCTCAA
GTACAGACCAGCCACACGCCACTGCGGACGGTTGGTGTTTTGCCTTAGCTACAGCGAGTT
N G I D V A Q>
A DOMAIN
H V W S V C G D A C Q P Q N G I D V A Q>
CODING REGION
H V W S V C G D A C Q P Q>
B DOMAIN
250 260 270 280 290 300
AAATGTTGCTCCACTGATTGCAGCTCCGATTACATCAAAGAAATCTGCTGCCCATTTGAC
TTTACAACGAGGTGACTAACGTCGAGGCTAATGTAGTTTCTTTAGACGACGGGTAAACTG
K C C S T D C S S D Y I K E I C C P F D>
A DOMAIN
K C C S T D C S S D Y I K E I C C P F D>
CODING REGION
TAA
ATT
*>
—>
*>
—>

FIG. 9B

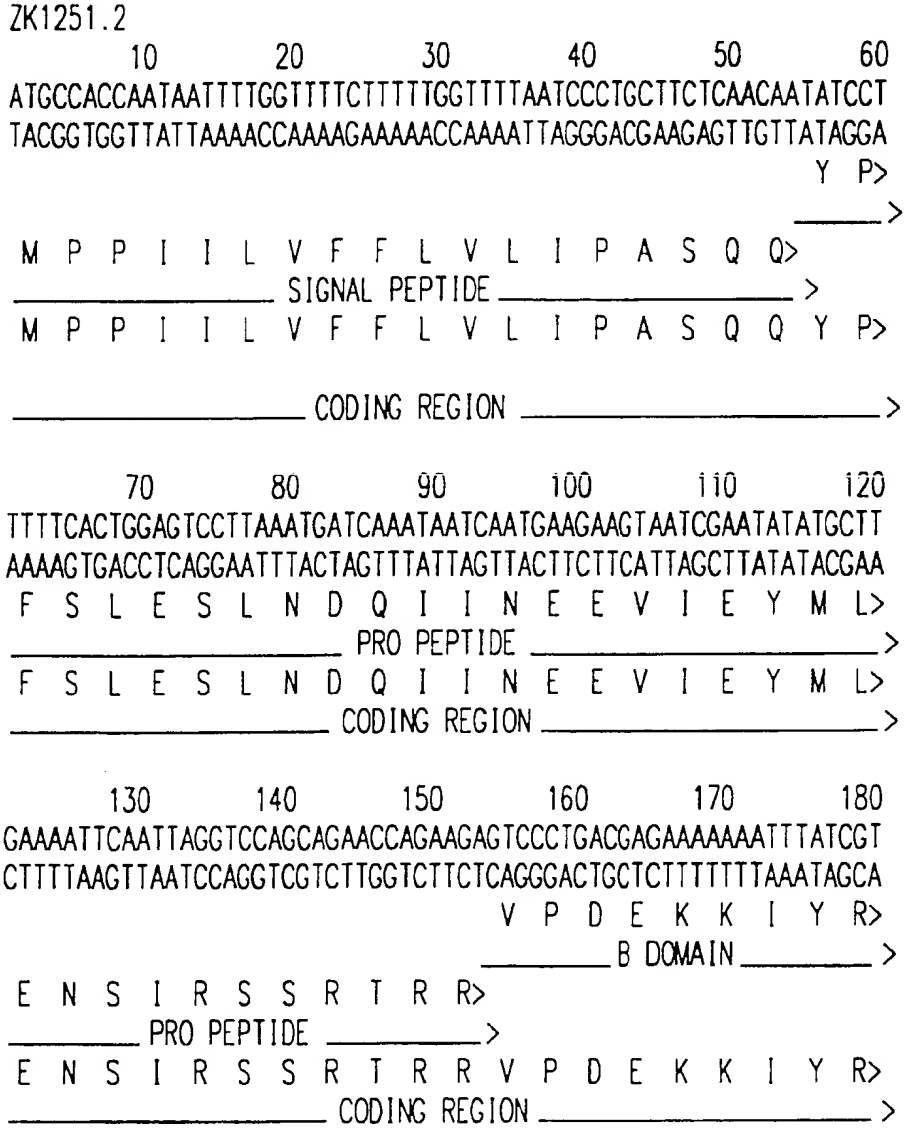

ZK1251.2
10 20 30 40 50 60
ATGCCACCAATAATTTTGGTTTTCTTTTTGGTTTTAATCCCTGCTTCTCAACAATATCCT
TACGGTGGTTATTAAAACCAAAAGAAAAACCAAAATTAGGGACGAAGAGTTGTTATAGGA
Y P>
M P P I I L V F F L V L I P A S Q Q>
SIGNAL PEPTIDE
M P P I I L V F F L V L I P A S Q Q Y P>
CODING REGION
70 80 90 100 110 120
TTTTCACTGGAGTCCTTAAATGATCAAATAATCAATGAAGAAGTAATCGAATATATGCTT
AAAAGTGACCTCAGGAATTTACTAGTTTATTAGTTACTTCTTCATTAGCTTATATACGAA
F S L E S L N D Q I I N E E V I E Y M L>
PRO PEPTIDE
F S L E S L N D Q I I N E E V I E Y M L>
CODING REGION
130 140 150 160 170 180
GAAAATTCAATTAGGTCCAGCAGAACCAGAAGAGTCCCTGACGAGAAAAAAATTTATCGT
CTTTTAAGTTAATCCAGGTCGTCTTGGTCTTCTCAGGGACTGCTCTTTTTTTAAATAGCA
V P D E K K I Y R>
B DOMAIN
E N S I R S S R T R R>
PRO PEPTIDE
E N S I R S S R T R R V P D E K K I Y R>
CODING REGION

FIG. 10A

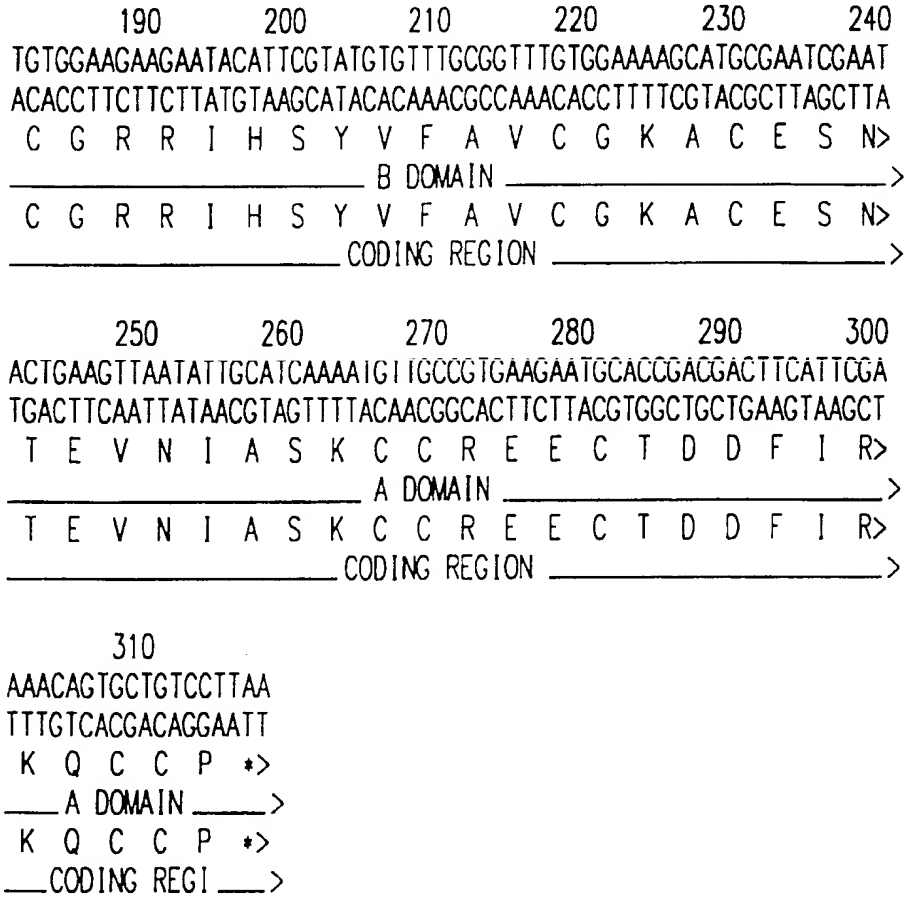

190 200 210 220 230 240
TGTGGAAGAAGAATACATTCGTATGTGTTTGCGGTTTGTGGAAAAGCATGCGAATCGAAT
ACACCTTCTTCTTATGTAAGCATACACAAACGCCAAACACCTTTTCGTACGCTTAGCTTA
C G R R I H S Y V F A V C G K A C E S N>
B DOMAIN
C G R R I H S Y V F A V C G K A C E S N>
CODING REGION
250 260 270 280 290 300
ACTGAAGTTAATATTGCATCAAAAIGTTGCCGTGAAGAATGCACCGACGACTTCATTCGA
TGACTTCAATTATAACGTAGTTTTACAACGGCACTTCTTACGTGGCTGCTGAAGTAAGCT
T E V N I A S K C C R E E C T D D F I R>
A DOMAIN
T E V N I A S K C C R E E C T D D F I R>
CODING REGION
310
AAACAGTGCTGTCCTTAA
TTTGTCACGACAGGAATT
K Q C C P *>
A DOMAIN
K Q C C P *>
CODING REGI

FIG. 10B

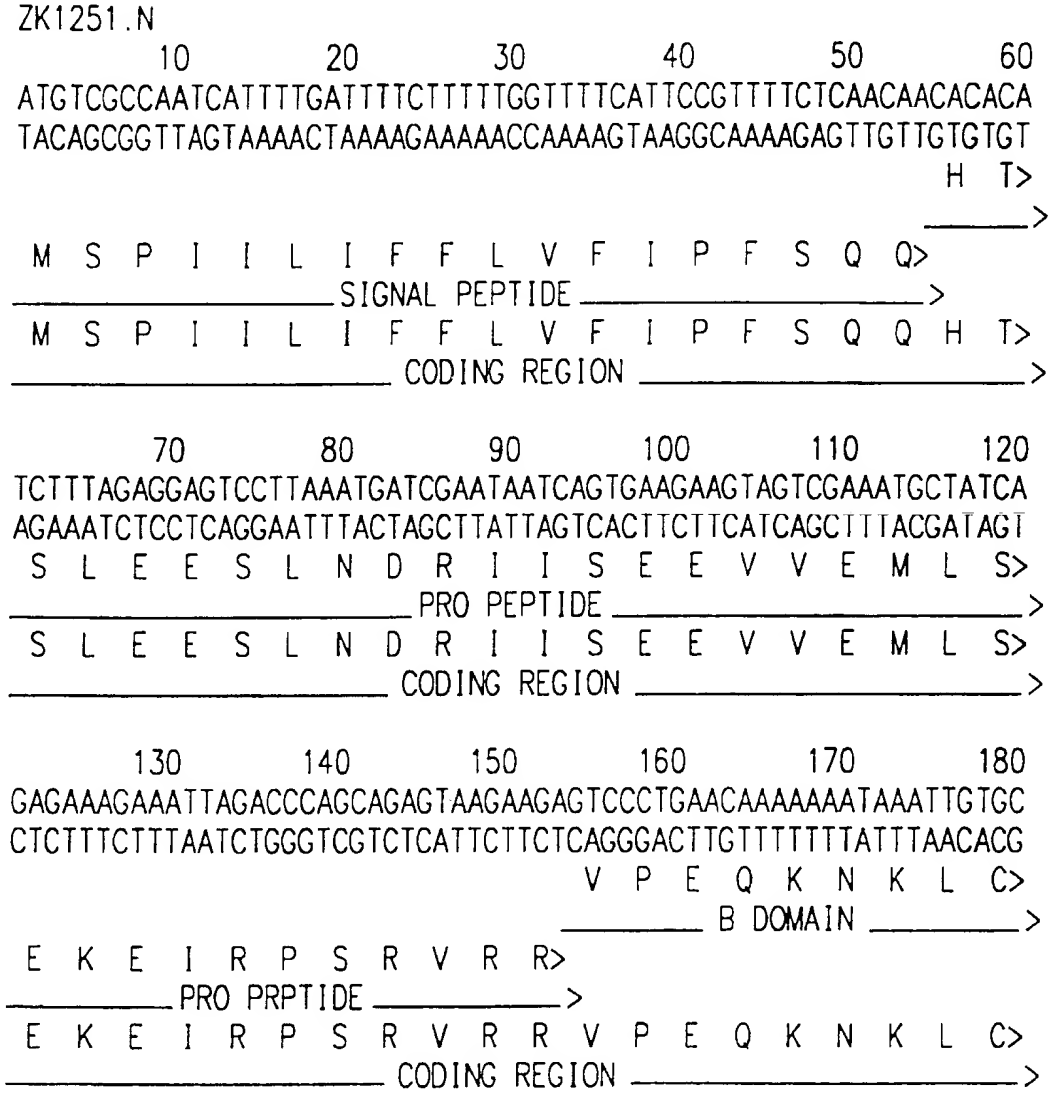
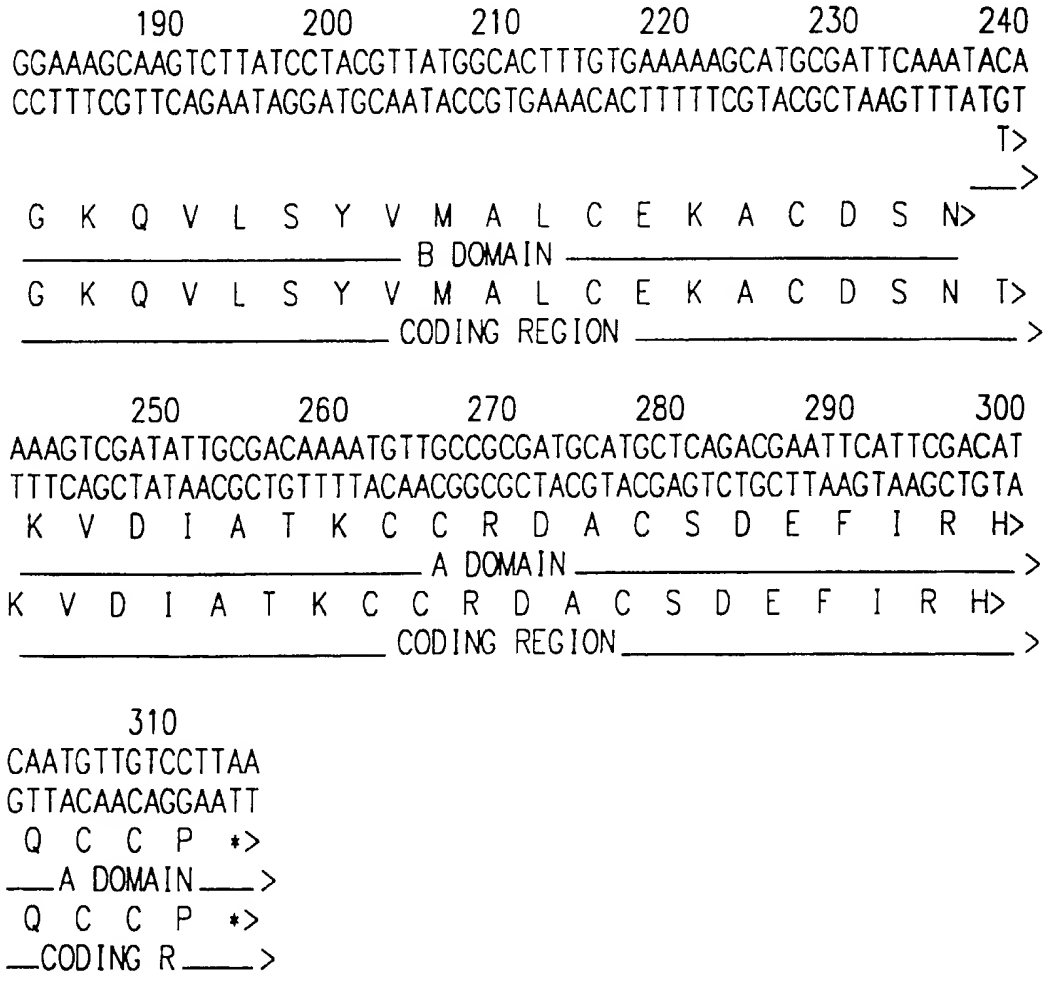

ZK1251.N
10 20 30 40 50 60
ATGTCGCCAATCATTTTGATTTTCTTTTTGGTTTTCATTCCGTTTTCTCAACAACACACA
TACAGCGGTTAGTAAAACTAAAAGAAAAACCAAAAGTAAGGCAAAAGAGTTGTTGTGTGT
H T>
M S P I I L I F F L V F I P F S Q Q>
SIGNAL PEPTIDE
M S P I I L I F F L V F I P F S Q Q H T>
CODING REGION
70 80 90 100 110 120
TCTTTAGAGGAGTCCTTAAATGATCGAATAATCAGTGAAGAAGTAGTCGAAATGCTATCA
AGAAATCTCCTCAGGAATTTACTAGCTTATTAGTCACTTCTTCATCAGCTTTACGATAGT
S L E E S L N D R I I S E E V V E M L S>
PRO PEPTIDE
S L E E S L N D R I I S E E V V E M L S>
CODING REGION
130 140 150 160 170 180
GAGAAAGAAATTAGACCCAGCAGAGTAAGAAGAGTCCCTGAACAAAAAAATAAATTGTGC
CTCTTTCTTTAATCTGGGTCGTCTCATTCTTCTCAGGGACTTGTTTTTTTATTTAACACG
V P E Q K N K L C>
B DOMAIN
E K E I R P S R V R R>
PRO PRPTIDE
E K E I R P S R V R R V P E Q K N K L C>
CODING REGION

FIG. 11A


190 200 210 220 230 240
GGAAAGCAAGTCTTATCCTACGTTATGGCACTTTGTGAAAAAGCATGCGATTCAAATACA
CCTTTCGTTCAGAATAGGATGCAATACCGTGAAACACTTTTTCGTACGCTAAGTTTATGT
T>
—>
G K Q V L S Y V M A L C E K A C D S N>
B DOMAIN
G K Q V L S Y V M A L C E K A C D S N T>
CODING REGION >
250 260 270 280 290 300
AAAGTCGATATTGCGACAAAATGTTGCCGCGATGCATGCTCAGACGAATTCATTCGACAT
TTTCAGCTATAACGCTGTTTTACAACGGCGCTACGTACGAGTCTGCTTAAGTAAGCTGTA
K V D I A T K C C R D A C S D E F I R H>
A DOMAIN >
K V D I A T K C C R D A C S D E F I R H>
CODING REGION >
310
CAATGTTGTCCTTAA
GTTACAACAGGAATT
Q C C P *>
A DOMAIN >
Q C C P *>
CODING R >
FIG. 11B

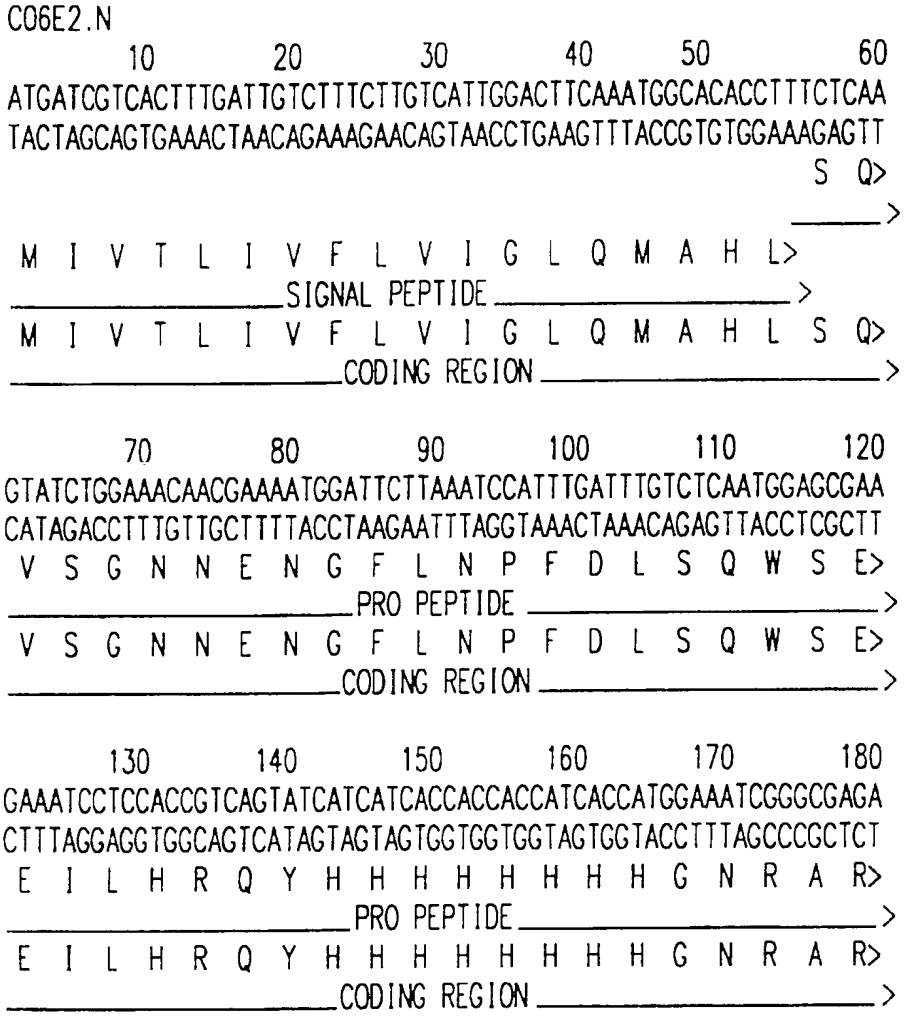

C06E2.N
10 20 30 40 50 60
ATGATCGTCACTTTGATTGTCTTTCTTGTCATTGGACTTCAAATGGCACACCTTTCTCAA
TACTAGCAGTGAAACTAACAGAAAGAACAGTAACCTGAAGTTTACCGTGTGGAAAGAGTT
S Q>
M I V T L I V F L V I G L Q M A H L>
SIGNAL PEPTIDE
M I V T L I V F L V I G L Q M A H L S Q>
CODING REGION
70 80 90 100 110 120
GTATCTGGAAACAACGAAAATGGATTCTTAAATCCATTTGATTTGTCTCAATGGAGCGAA
CATAGACCTTTGTTGCTTTTACCTAAGAATTTAGGTAAACTAAACAGAGTTACCTCGCTT
V S G N N E N G F L N P F D L S Q W S E>
PRO PEPTIDE
V S G N N E N G F L N P F D L S Q W S E>
CODING REGION
130 140 150 160 170 180
GAAATCCTCCACCGTCAGTATCATCATCACCACCACCATCACCATGGAAATCGGGCGAGA
CTTTAGGAGGTGGCAGTCATAGTAGTAGTGGTGGTGGTAGTGGTACCTTTAGCCCGCTCT
E I L H R Q Y H H H H H H H H H G N R A R>
PRO PEPTIDE
E I L H R Q Y H H H H H H H H H G N R A R>
CODING REGION

FIG. 12A

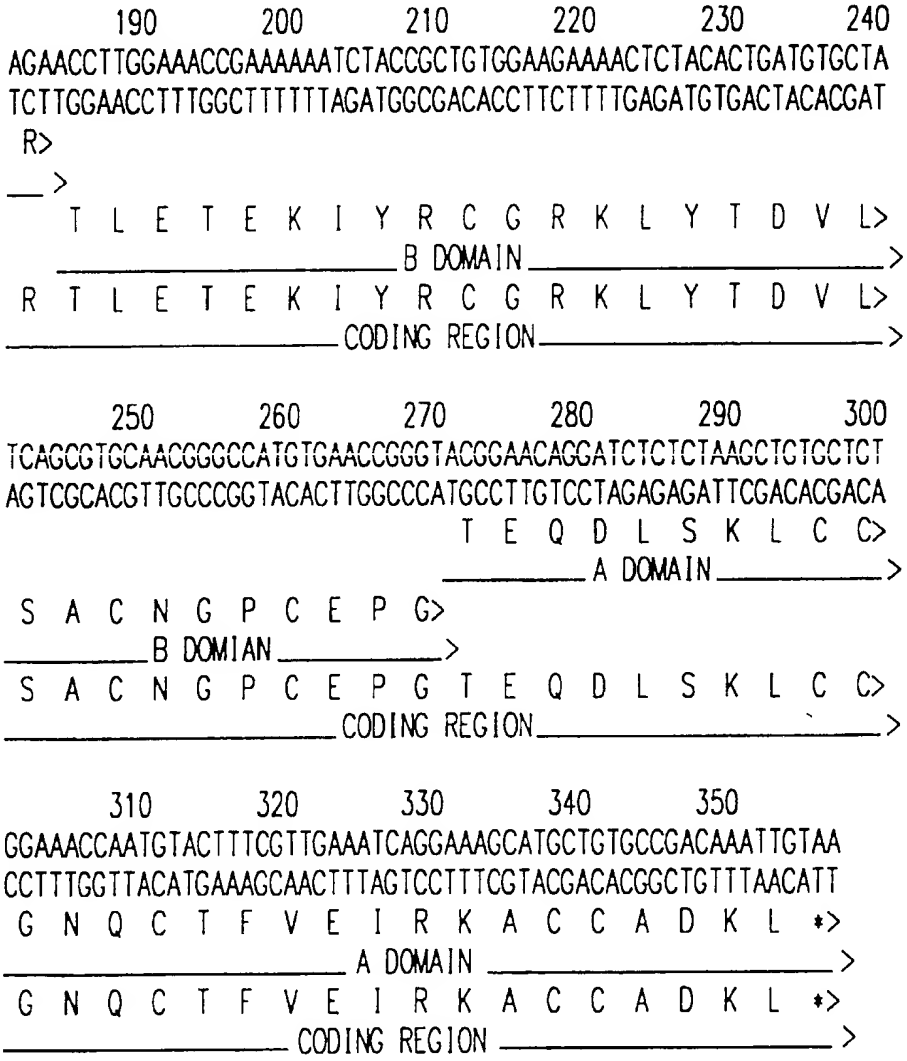

190 200 210 220 230 240
AGAACCTTGGAAACCGAAAAAATCTACCGCTGTGGAAGAAAACTCTACACTGATGTGCTA
TCTTGGAACCTTTGGCTTTTTTAGATGGCGACACCTTCTTTTGAGATGTGACTACACGAT
R>
>
T L E T E K I Y R C G R K L Y T D V L>
B DOMAIN
R T L E T E K I Y R C G R K L Y T D V L>
CODING REGION
250 260 270 280 290 300
TCAGCGTGCAACGGGCCATGTGAACCGGGTACGGAACAGGATCTCTCTAAGCTGTGCTCT
AGTCGCACGTTGCCCGGTACACTTGGCCCATGCCTTGTCCTAGAGAGATTCGACACGACA
T E Q D L S K L C C>
A DOMAIN
S A C N G P C E P G>
B DOMIAN
S A C N G P C E P G T E Q D L S K L C C>
CODING REGION
310 320 330 340 350
GGAAACCAATGTACTTTCGTTGAAATCAGGAAAGCATGCTGTGCCGACAAATTGTAA
CCTTTGGTTACATGAAAGCAACTTTAGTCCTTTCGTACGACACGGCTGTTTAACATT
G N Q C T F V E I R K A C C A D K L *>
A DOMAIN
G N Q C T F V E I R K A C C A D K L *>
CODING REGION

FIG. 12B

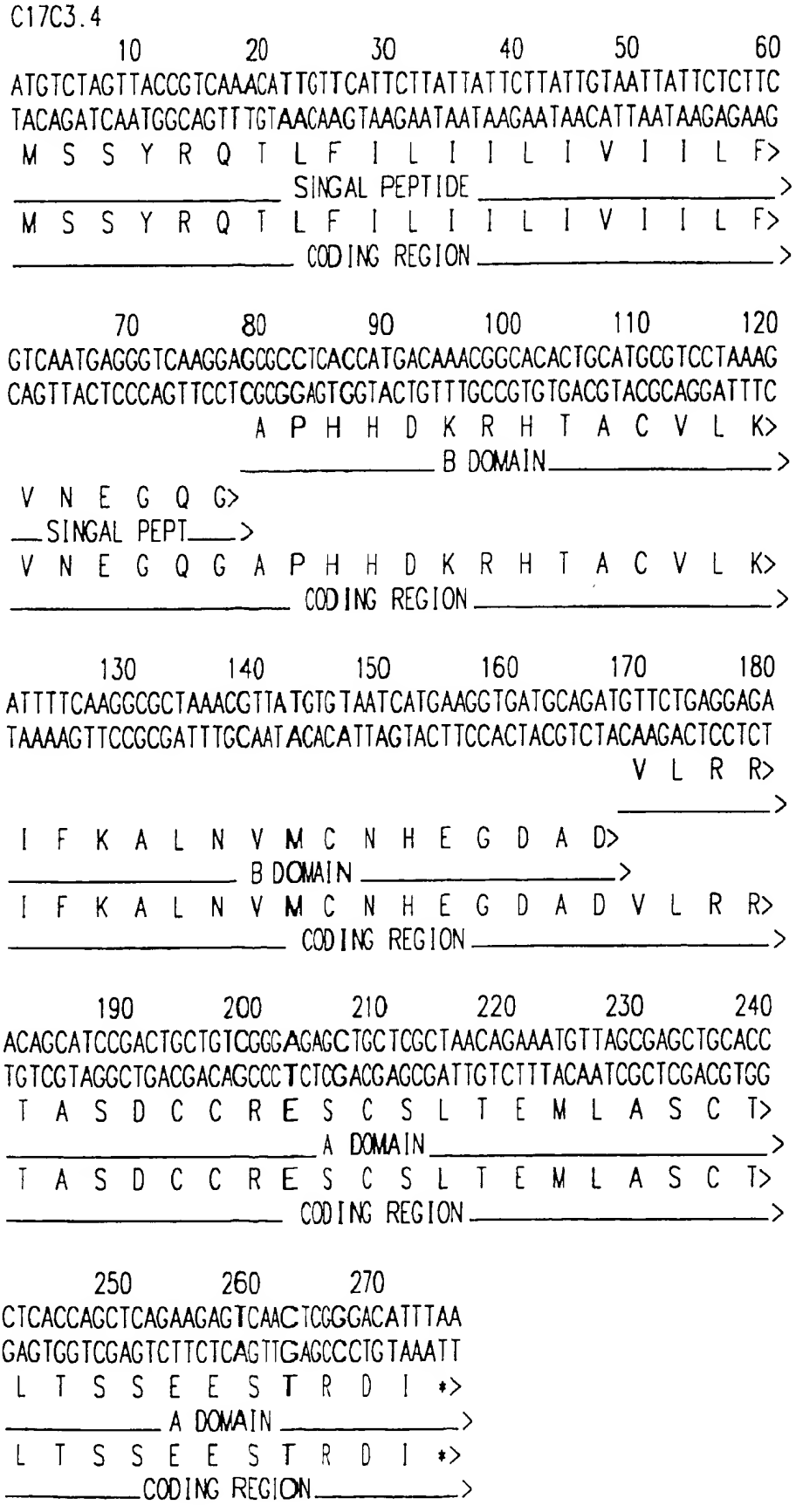

C17C3.4
10 20 30 40 50 60
ATGTCTAGTTACCGTCAAACATTGTTCATTCTTATTATTCTTATTGTAATTATTCTCTTC
TACAGATCAATGGCAGTTTGTAACAAGTAAGAATAATAAGAATAACATTAATAAGAGAAG
M S S Y R Q T L F I L I I L I V I I L F>
SINGAL PEPTIDE
M S S Y R Q T L F I L I I L I V I I L F>
CODING REGION
70 80 90 100 110 120
GTCAATGAGGGTCAAGGAGCGCCTCACCATGACAAACGGCACACTGCATGCGTCCTAAAG
CAGTTACTCCCAGTTCCTCGCGGAGTGGTACTGTTTGCCGTGTGACGTACGCAGGATTTC
A P H H D K R H T A C V L K>
B DOMAIN
V N E G Q G>
SINGAL PEPT
V N E G Q G A P H H D K R H T A C V L K>
CODING REGION
130 140 150 160 170 180
ATTTTCAAGGCGCTAAACGTTATGTGTAATCATGAAGGTGATGCAGATGTTCTGAGGAGA
TAAAAGTTCCGCGATTTGCAATACACATTAGTACTTCCACTACGTCTACAAGACTCCTCT
V L R R>
I F K A L N V M C N H E G D A D>
B DOMAIN
I F K A L N V M C N H E G D A D V L R R>
CODING REGION
190 200 210 220 230 240
ACAGCATCCGACTGCTGTCGGGAGAGCTGCTCGCTAACAGAAATGTTAGCGAGCTGCACC
TGTCGTAGGCTGACGACAGCCCTCTCGACGAGCGATTGTCTTTACAATCGCTCGACGTGG
T A S D C C R E S C S L T E M L A S C T>
A DOMAIN
T A S D C C R E S C S L T E M L A S C T>
CODING REGION
250 260 270
CTCACCAGCTCAGAAGAGTCAACTCGGGACATTTAA
GAGTGGTCGAGTCTTCTCAGTTGAGCCCTGTAAATT
L T S S E E S T R D I *>
A DOMAIN
L T S S E E S T R D I *>
CODING REGION

FIG. 13

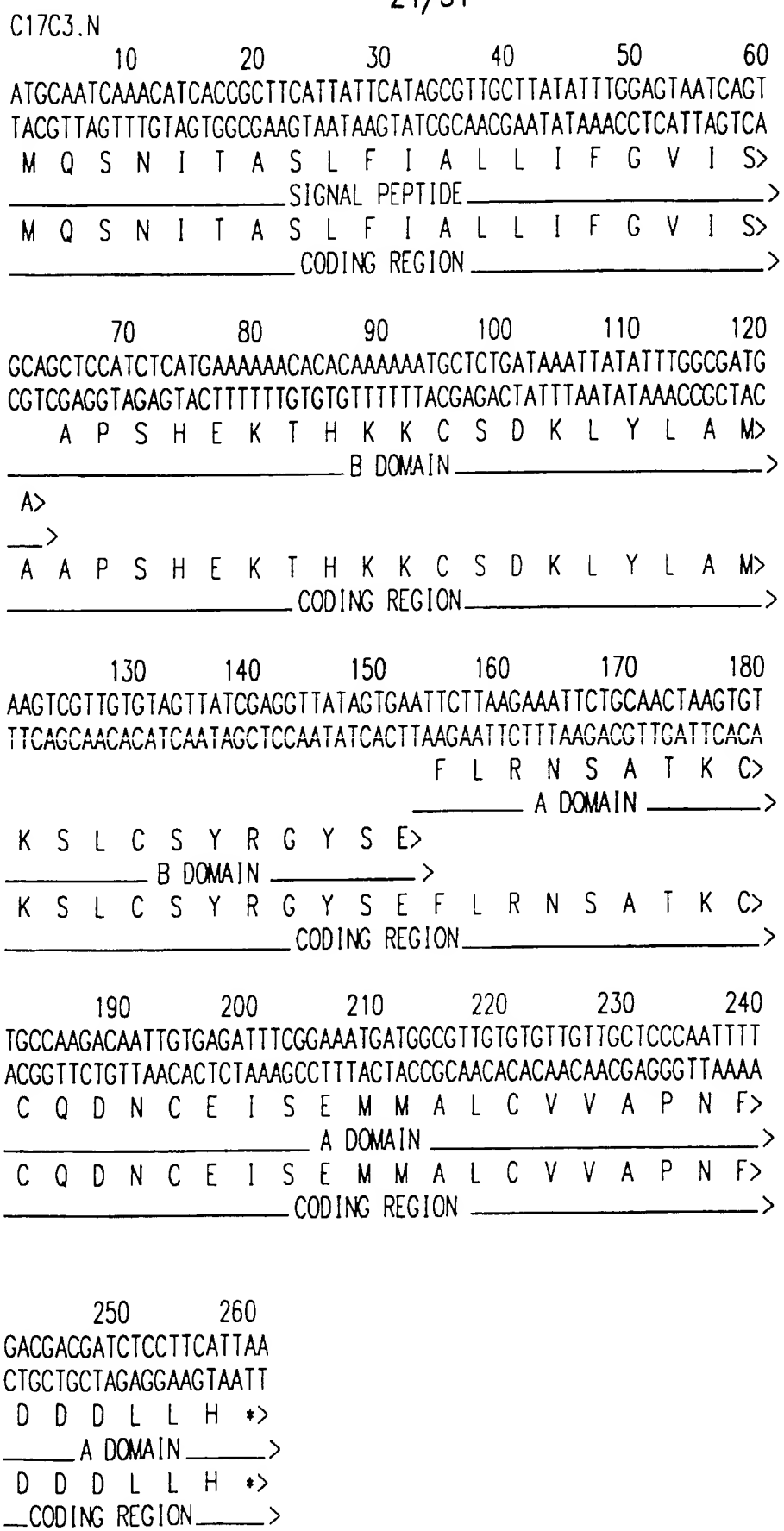

C17C3.N
10 20 30 40 50 60
ATGCAATCAAACATCACCGCTTCATTATTCATAGCGTTGCTTATATTTGGAGTAATCAGT
TACGTTAGTTTGTAGTGGCGAAGTAATAAGTATCGCAACGAATATAAACCTCATTAGTCA
M Q S N I T A S L F I A L L I F G V I S>
SIGNAL PEPTIDE
M Q S N I T A S L F I A L L I F G V I S>
CODING REGION
70 80 90 100 110 120
GCAGCTCCATCTCATGAAAAAACACACAAAAAATGCTCTGATAAATTATATTTGGCGATG
CGTCGAGGTAGAGTACTTTTTTGTGTGTTTTTTACGAGACTATTTAATATAAACCGCTAC
A P S H E K T H K K C S D K L Y L A M>
B DOMAIN
A>
A A P S H E K T H K K C S D K L Y L A M>
CODING REGION
130 140 150 160 170 180
AAGTCGTTGTGTAGTTATCGAGGTTATAGTGAATTCTTAAGAAATTCTGCAACTAAGTGT
TTCAGCAACACATCAATAGCTCCAATATCACTTAAGAATTCTTTAAGACGTTGATTCACA
F L R N S A T K C>
A DOMAIN
K S L C S Y R G Y S E>
B DOMAIN
K S L C S Y R G Y S E F L R N S A T K C>
CODING REGION
190 200 210 220 230 240
TGCCAAGACAATTGTGAGATTTCGGAAATGATGGCGTTGTGTGTTGTTGCTCCCAATTTT
ACGGTTCTGTTAACACTCTAAAGCCTTTACTACCGCAACACACAACAACGAGGGTTAAAA
C Q D N C E I S E M M A L C V V A P N F>
A DOMAIN
C Q D N C E I S E M M A L C V V A P N F>
CODING REGION
250 260
GACGACGATCTCCTTCATTAA
CTGCTGCTAGAGGAAGTAATT
D D D L L H *>
A DOMAIN
D D D L L H *>
CODING REGION

FIG. 14

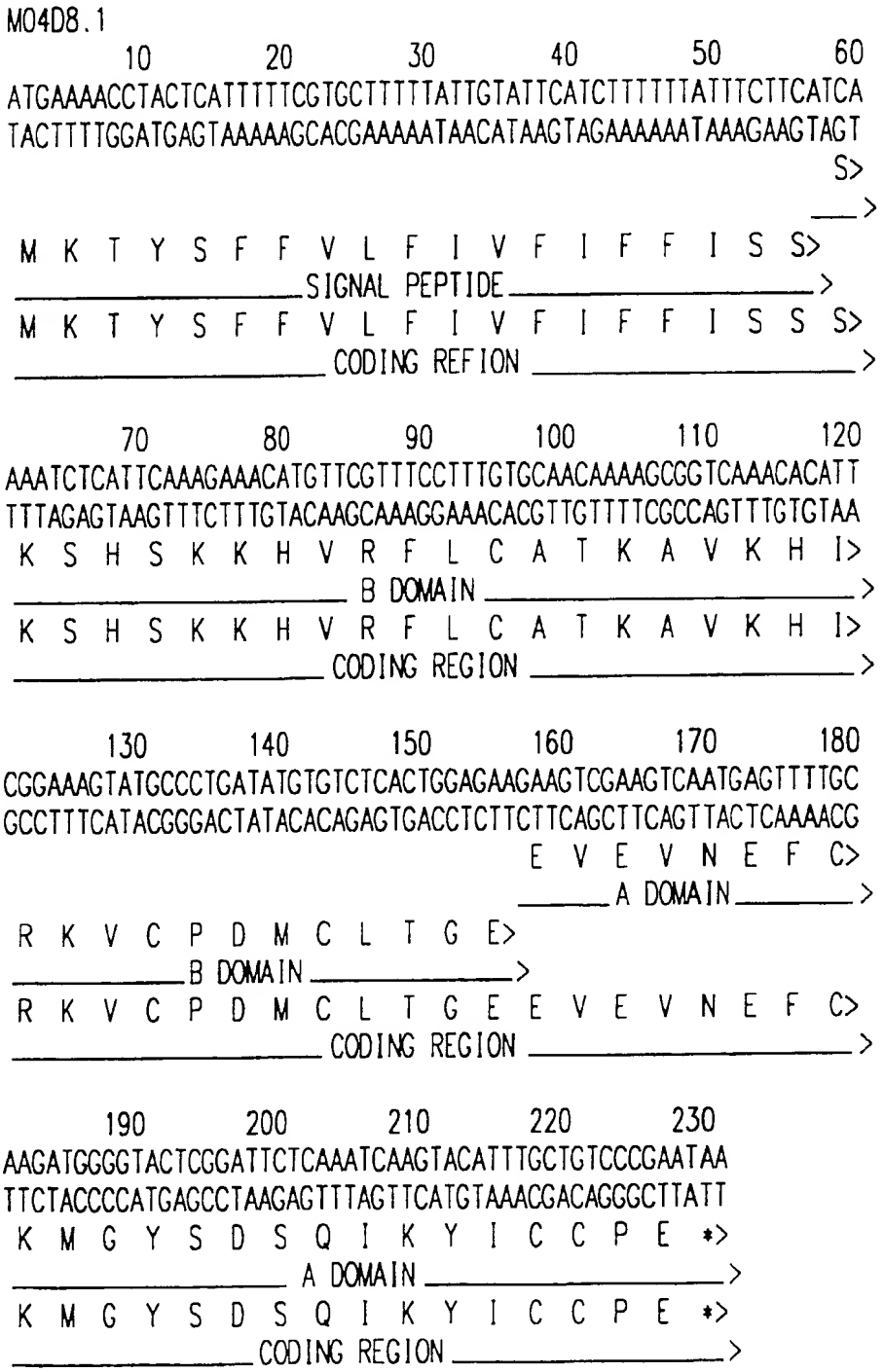

M04D8.1
10 20 30 40 50 60
ATGAAAACCTACTCATTTTTCGTGCTTTTTATTGTATTCATCTTTTTTATTTCTTCATCA
TACTTTTGGATGAGTAAAAAGCACGAAAAATAACATAAGTAGAAAAAATAAAGAAGTAGT
S>
M K T Y S F F V L F I V F I F F I S S>
SIGNAL PEPTIDE
M K T Y S F F V L F I V F I F F I S S S>
CODING REFION
70 80 90 100 110 120
AAATCTCATTCAAAGAAACATGTTCGTTTCCTTTGTGCAACAAAAGCGGTCAAACACATT
TTTAGAGTAAGTTTCTTTGTACAAGCAAAGGAAACACGTTGTTTTCGCCAGTTTGTGTAA
K S H S K K H V R F L C A T K A V K H I>
B DOMAIN
K S H S K K H V R F L C A T K A V K H I>
CODING REGION
130 140 150 160 170 180
CGGAAAGTATGCCCTGATATGTGTCTCACTGGAGAAGAAGTCGAAGTCAATGAGTTTTGC
GCCTTTCATACGGGACTATACACAGAGTGACCTCTTCTTCAGCTTCAGTTACTCAAAACG
E V E V N E F C>
A DOMAIN
R K V C P D M C L T G E>
B DOMAIN
R K V C P D M C L T G E E V E V N E F C>
CODING REGION
190 200 210 220 230
AAGATGGGGTACTCGGATTCTCAAATCAAGTACATTTGCTGTCCCGAATAA
TTCTACCCCATGAGCCTAAGAGTTTAGTTCATGTAAACGACAGGGCTTATT
K M G Y S D S Q I K Y I C C P E *>
A DOMAIN
K M G Y S D S Q I K Y I C C P E *>
CODING REGION

FIG. 15

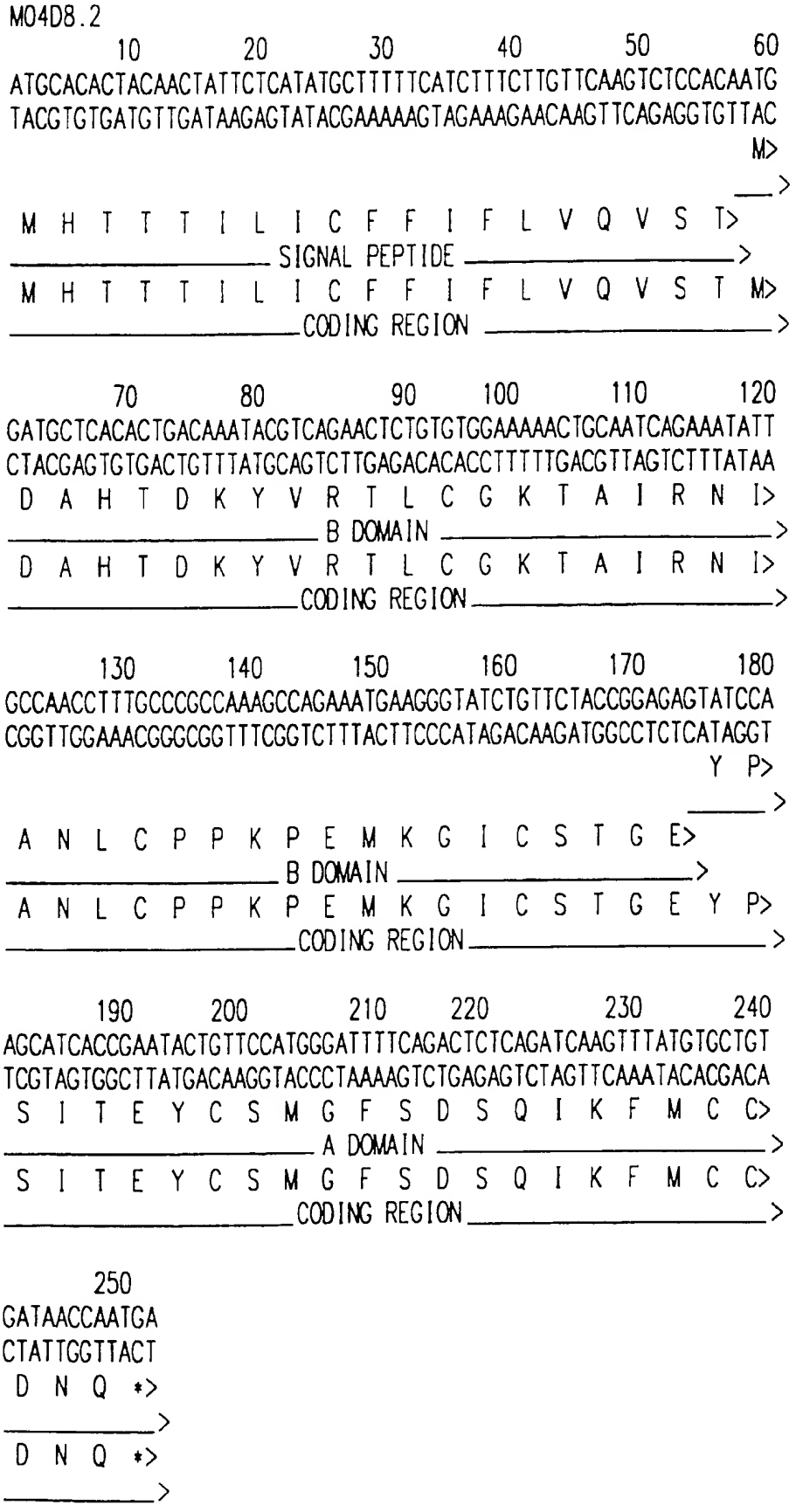

M04D8.2
10 20 30 40 50 60
ATGCACACTACAACTATTCTCATATGCTTTTTCATCTTTCTTGTTCAAGTCTCCACAATG
TACGTGTGATGTTGATAAGAGTATACGAAAAAGTAGAAAGAACAAGTTCAGAGGTGTTAC
M>
M H T T T I L I C F F I F L V Q V S T>
SIGNAL PEPTIDE
M H T T T I L I C F F I F L V Q V S T M>
CODING REGION
70 80 90 100 110 120
GATGCTCACACTGACAAATACGTCAGAACTCTGTGTGGAAAAACTGCAATCAGAAATATT
CTACGAGTGTGACTGTTTATGCAGTCTTGAGACACACCTTTTTGACGTTAGTCTTTATAA
D A H T D K Y V R T L C G K T A I R N I>
B DOMAIN
D A H T D K Y V R T L C G K T A I R N I>
CODING REGION
130 140 150 160 170 180
GCCAACCTTTGCCCGCCAAAGCCAGAAATGAAGGGTATCTGTTCTACCGGAGAGTATCCA
CGGTTGGAAACGGGCGGTTTCGGTCTTTACTTCCCATAGACAAGATGGCCTCTCATAGGT
Y P>
A N L C P P K P E M K G I C S T G E>
B DOMAIN
A N L C P P K P E M K G I C S T G E Y P>
CODING REGION
190 200 210 220 230 240
AGCATCACCGAATACTGTTCCATGGGATTTTCAGACTCTCAGATCAAGTTTATGTGCTGT
TCGTAGTGGCTTATGACAAGGTACCCTAAAAGTCTGAGAGTCTAGTTCAAATACACGACA
S I T E Y C S M G F S D S Q I K F M C C>
A DOMAIN
S I T E Y C S M G F S D S Q I K F M C C>
CODING REGION
250
GATAACCAATGA
CTATTGGTTACT
D N Q *>
D N Q *>

FIG. 16

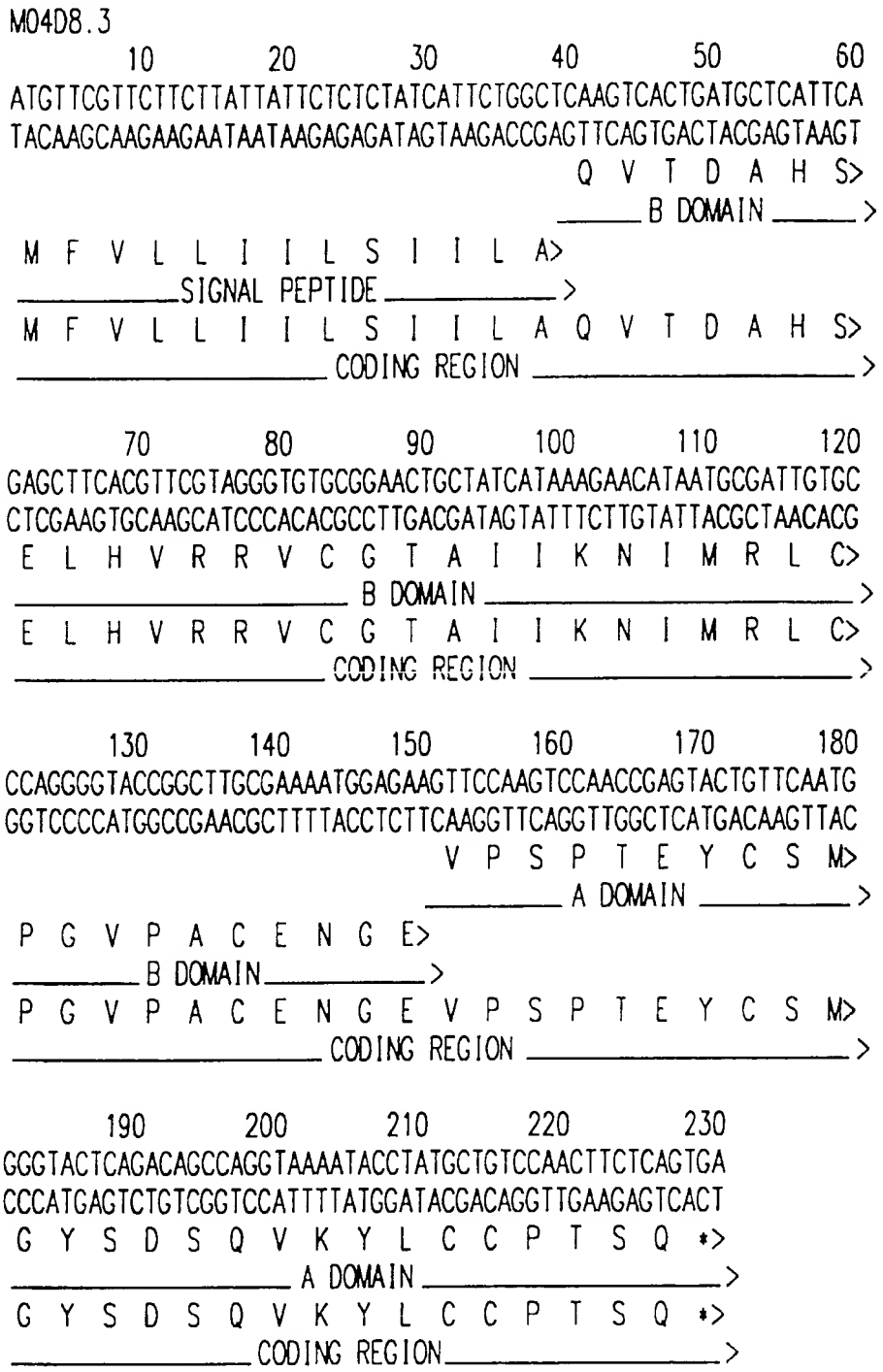

M04D8.3
10 20 30 40 50 60
ATGTTCGTTCTTCTTATTATTCTCTCTATCATTCTGGCTCAAGTCACTGATGCTCATTCA
TACAAGCAAGAAGAATAATAAGAGAGATAGTAAGACCGAGTTCAGTGACTACGAGTAAGT
Q V T D A H S>
B DOMAIN >
M F V L L I I L S I I L A>
SIGNAL PEPTIDE >
M F V L L I I L S I I L A Q V T D A H S>
CODING REGION >
70 80 90 100 110 120
GAGCTTCACGTTCGTAGGGTGTGCGGAACTGCTATCATAAAGAACATAATGCGATTGTGC
CTCGAAGTGCAAGCATCCCACACGCCTTGACGATAGTATTTCTTGTATTACGCTAACACG
E L H V R R V C G T A I I K N I M R L C>
B DOMAIN >
E L H V R R V C G T A I I K N I M R L C>
CODING REGION >
130 140 150 160 170 180
CCAGGGGTACCGGCTTGCGAAAATGGAGAAGTTCCAAGTCCAACCGAGTACTGTTCAATG
GGTCCCCATGGCCGAACGCTTTTACCTCTTCAAGGTTCAGGTTGGCTCATGACAAGTTAC
V P S P T E Y C S M>
A DOMAIN >
P G V P A C E N G E>
B DOMAIN >
P G V P A C E N G E V P S P T E Y C S M>
CODING REGION >
190 200 210 220 230
GGGTACTCAGACAGCCAGGTAAAATACCTATGCTGTCCAACTTCTCAGTGA
CCCATGAGTCTGTCGGTCCATTTTATGGATACGACAGGTTGAAGAGTCACT
G Y S D S Q V K Y L C C P T S Q *>
A DOMAIN >
G Y S D S Q V K Y L C C P T S Q *>
CODING REGION>

FIG. 17

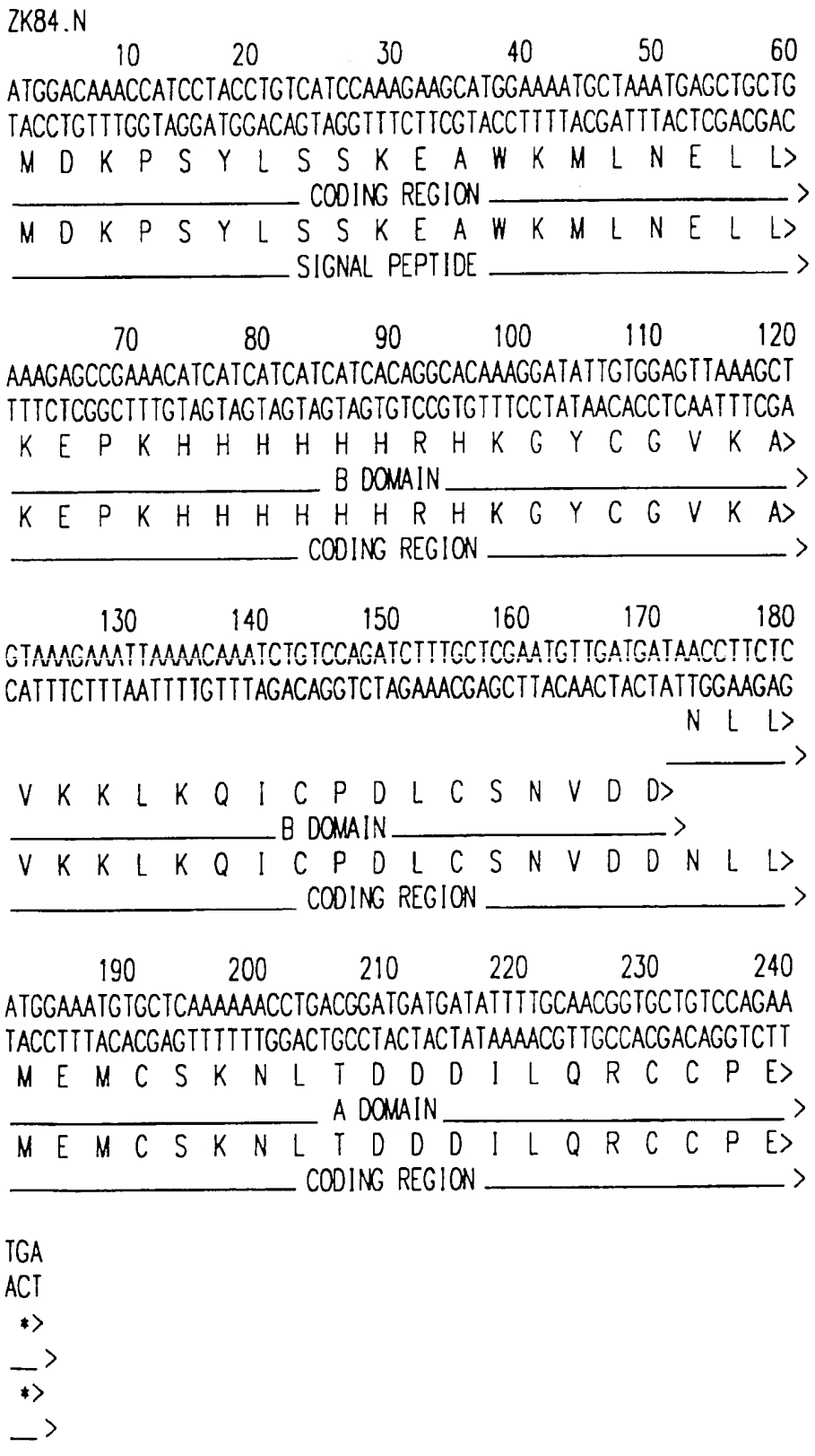

ZK84.N
10 20 30 40 50 60
ATGGACAAACCATCCTACCTGTCATCCAAAGAAGCATGGAAAATGCTAAATGAGCTGCTG
TACCTGTTTGGTAGGATGGACAGTAGGTTTCTTCGTACCTTTTACGATTTACTCGACGAC
M D K P S Y L S S K E A W K M L N E L L>
CODING REGION
M D K P S Y L S S K E A W K M L N E L L>
SIGNAL PEPTIDE
70 80 90 100 110 120
AAAGAGCCGAAACATCATCATCATCATCACAGGCACAAAGGATATTGTGGAGTTAAAGCT
TTTCTCGGCTTTGTAGTAGTAGTAGTAGTGTCCGTGTTTCCTATAACACCTCAATTTCGA
K E P K H H H H H H H R H K G Y C G V K A>
B DOMAIN
K E P K H H H H H H H R H K G Y C G V K A>
CODING REGION
130 140 150 160 170 180
GTAAAGAAATTAAAACAAATCTGTCCAGATCTTTGCTCGAATGTTGATGATAACCTTCTC
CATTTCTTTAATTTTGTTTAGACAGGTCTAGAAACGAGCTTACAACTACTATTGGAAGAG
N L L>
V K K L K Q I C P D L C S N V D D>
B DOMAIN
V K K L K Q I C P D L C S N V D D N L L>
CODING REGION
190 200 210 220 230 240
ATGGAAATGTGCTCAAAAAACCTGACGGATGATGATATTTTGCAACGGTGCTGTCCAGAA
TACCTTTACACGAGTTTTTTGGACTGCCTACTACTATAAAACGTTGCCACGACAGGTCTT
M E M C S K N L T D D D I L Q R C C P E>
A DOMAIN
M E M C S K N L T D D D I L Q R C C P E>
CODING REGION
TGA
ACT
*>
*>

FIG. 18

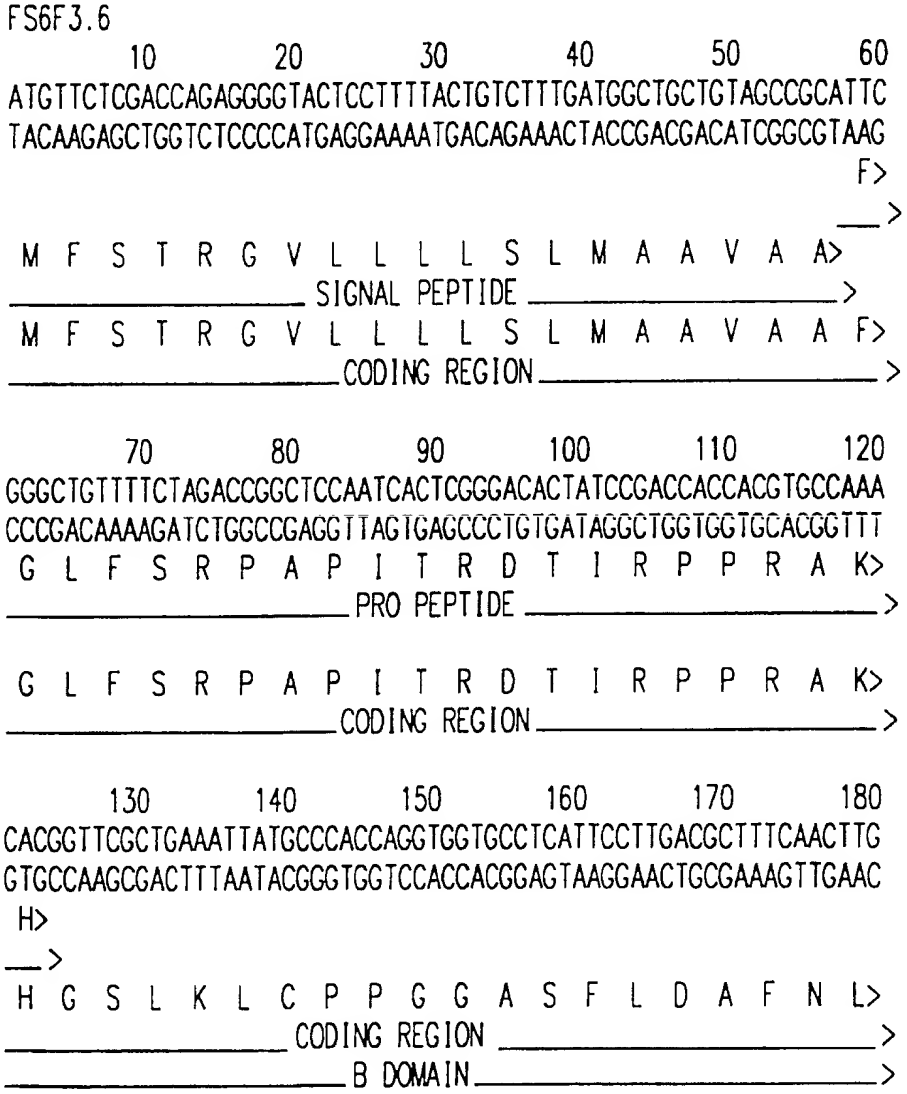

FS6F3.6
10 20 30 40 50 60
ATGTTCTCGACCAGAGGGGTACTCCTTTTACTGTCTTTGATGGCTGCTGTAGCCGCATTC
TACAAGAGCTGGTCTCCCCATGAGGAAAATGACAGAAACTACCGACGACATCGGCGTAAG
F>
M F S T R G V L L L L S L M A A V A A>
SIGNAL PEPTIDE
M F S T R G V L L L L S L M A A V A A F>
CODING REGION
70 80 90 100 110 120
GGGCTGTTTTCTAGACCGGCTCCAATCACTCGGGACACTATCCGACCACCACGTGCCAAA
CCCGACAAAAGATCTGGCCGAGGTTAGTGAGCCCTGTGATAGGCTGGTGGTGCACGGTTT
G L F S R P A P I T R D T I R P P R A K>
PRO PEPTIDE
G L F S R P A P I T R D T I R P P R A K>
CODING REGION
130 140 150 160 170 180
CACGGTTCGCTGAAATTATGCCCACCAGGTGGTGCCTCATTCCTTGACGCTTTCAACTTG
GTGCCAAGCGACTTTAATACGGGTGGTCCACCACGGAGTAAGGAACTGCGAAAGTTGAAC
H>
H G S L K L C P P G G A S F L D A F N L>
CODING REGION
B DOMAIN

FIG. 19A

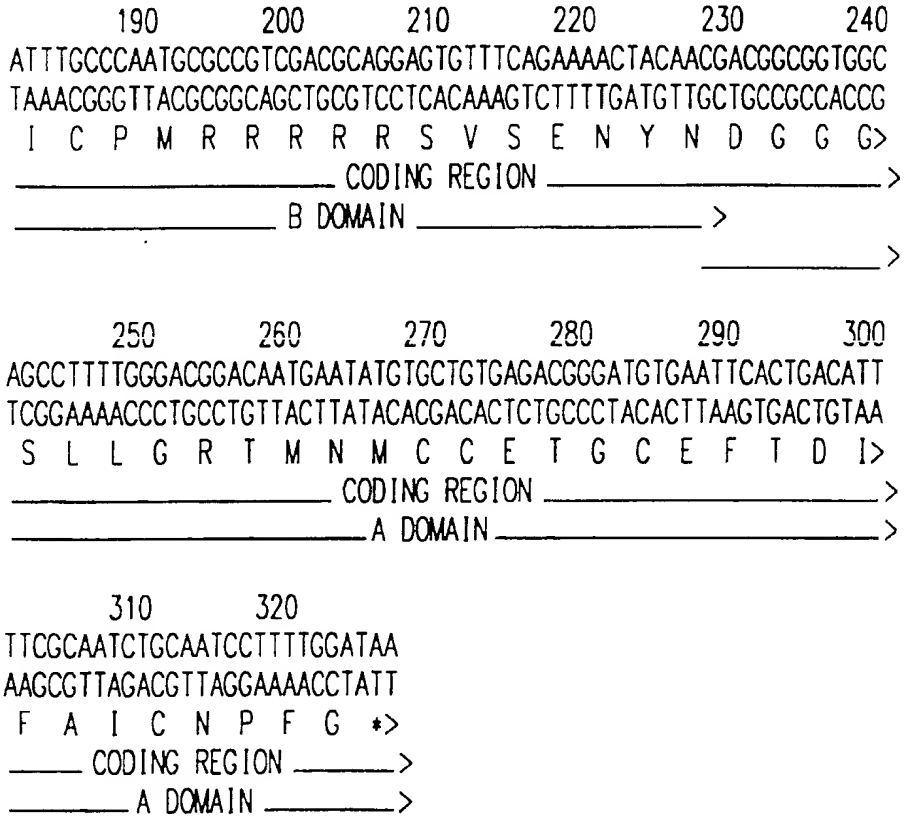

190 200 210 220 230 240
ATTTGCCCAATGCGCCGTCGACGCAGGAGTGTTTCAGAAAACTACAACGACGGCGGTGGC
TAAACGGGTTACGCGGCAGCTGCGTCCTCACAAAGTCTTTTGATGTTGCTGCCGCCACCG
I C P M R R R R R S V S E N Y N D G G G>
CODING REGION
B DOMAIN
250 260 270 280 290 300
AGCCTTTTGGGACGGACAATGAATATGTGCTGTGAGACGGGATGTGAATTCACTGACATT
TCGGAAAACCCTGCCTGTTACTTATACACGACACTCTGCCCTACACTTAAGTGACTGTAA
S L L G R T M N M C C E T G C E F T D I>
CODING REGION
A DOMAIN
310 320
TTCGCAATCTGCAATCCTTTTGGATAA
AAGCGTTAGACGTTAGGAAAACCTATT
F A I C N P F G *>
CODING REGION
A DOMAIN

FIG. 19B

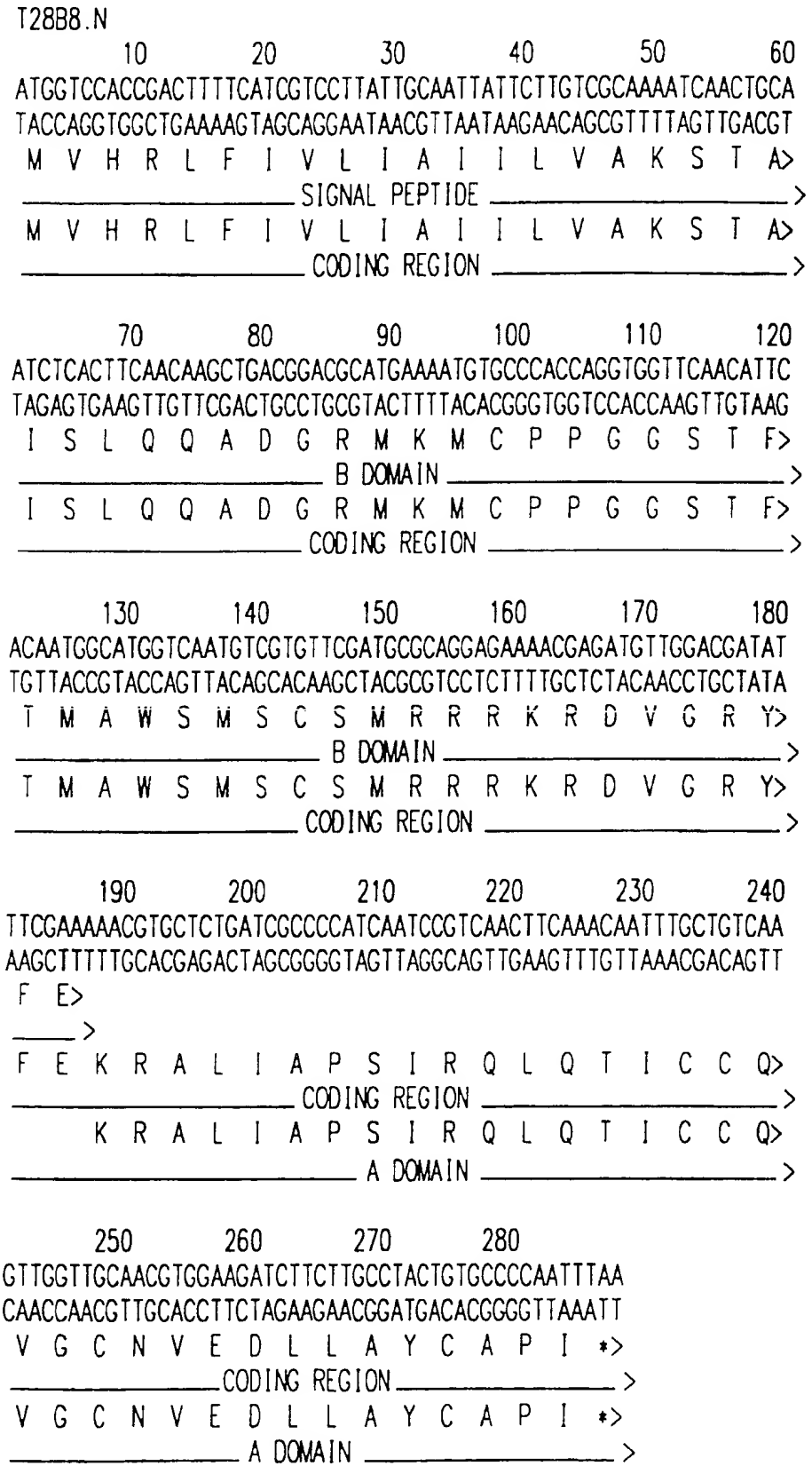

T28B8.N
10 20 30 40 50 60
ATGGTCCACCGACTTTTCATCGTCCTTATTGCAATTATTCTTGTCGCAAAATCAACTGCA
TACCAGGTGGCTGAAAAGTAGCAGGAATAACGTTAATAAGAACAGCGTTTTAGTTGACGT
M V H R L F I V L I A I I L V A K S T A>
SIGNAL PEPTIDE
M V H R L F I V L I A I I L V A K S T A>
CODING REGION
70 80 90 100 110 120
ATCTCACTTCAACAAGCTGACGGACGCATGAAAATGTGCCCACCAGGTGGTTCAACATTC
TAGAGTGAAGTTGTTCGACTGCCTGCGTACTTTTACACGGGTGGTCCACCAAGTTGTAAG
I S L Q Q A D G R M K M C P P G G S T F>
B DOMAIN
I S L Q Q A D G R M K M C P P G G S T F>
CODING REGION
130 140 150 160 170 180
ACAATGGCATGGTCAATGTCGTGTTCGATGCGCAGGAGAAAACGAGATGTTGGACGATAT
TGTTACCGTACCAGTTACAGCACAAGCTACGCGTCCTCTTTTGCTCTACAACCTGCTATA
T M A W S M S C S M R R R K R D V G R Y>
B DOMAIN
T M A W S M S C S M R R R K R D V G R Y>
CODING REGION
190 200 210 220 230 240
TTCGAAAAACGTGCTCTGATCGCCCCATCAATCCGTCAACTTCAAACAATTTGCTGTCAA
AAGCTTTTTGCACGAGACTAGCGGGGTAGTTAGGCAGTTGAAGTTTGTTAAACGACAGTT
F E>
F E K R A L I A P S I R Q L Q T I C C Q>
CODING REGION
K R A L I A P S I R Q L Q T I C C Q>
A DOMAIN
250 260 270 280
GTTGGTTGCAACGTGGAAGATCTTCTTGCCTACTGTGCCCCAATTTAA
CAACCAACGTTGCACCTTCTAGAAGAACGGATGACACGGGGTTAAATT
V G C N V E D L L A Y C A P I *>
CODING REGION
V G C N V E D L L A Y C A P I *>
A DOMAIN

FIG. 20

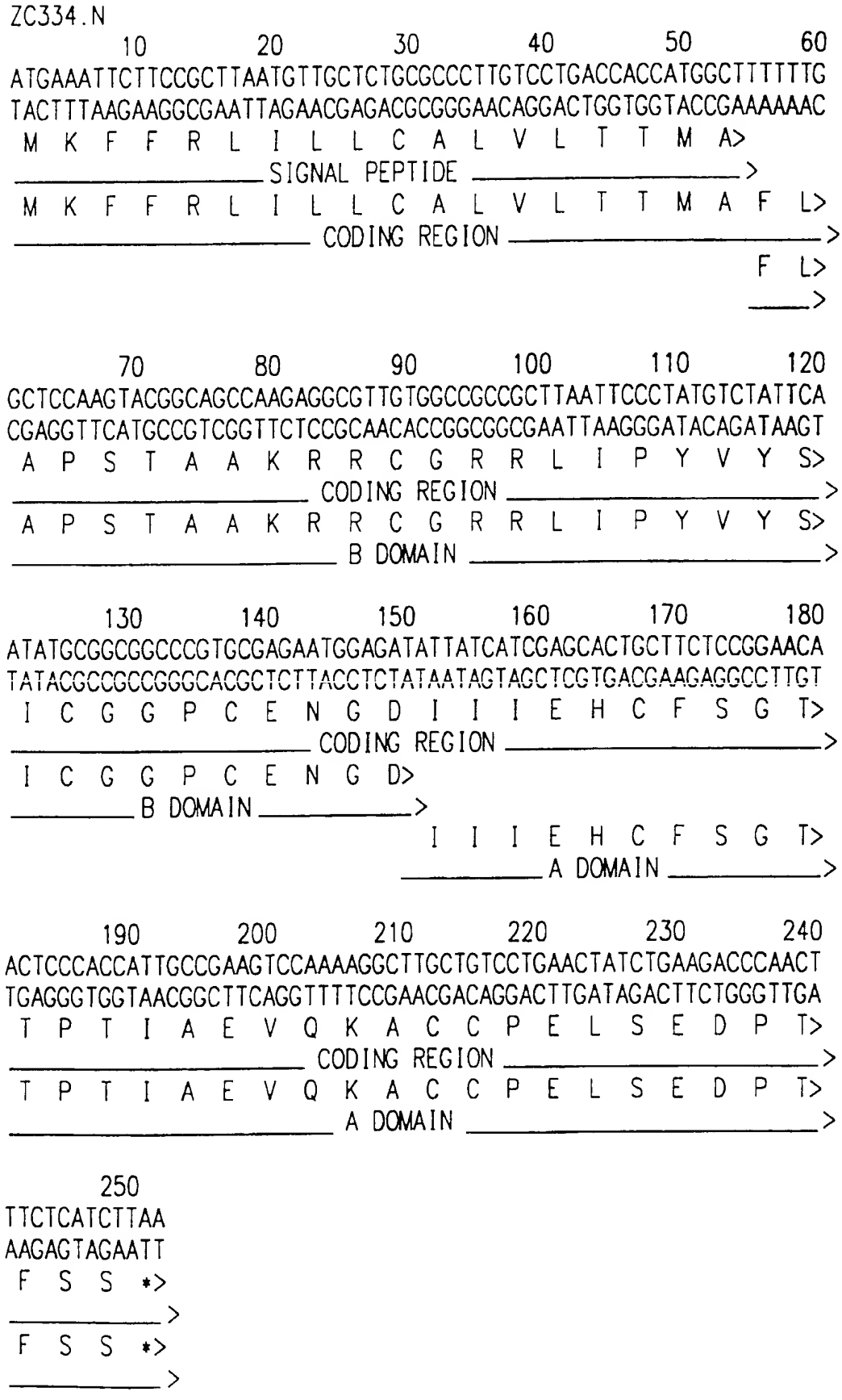

ZC334.N
10 20 30 40 50 60
ATGAAATTCTTCCGCTTAATGTTGCTCTGCGCCCTTGTCCTGACCACCATGGCTTTTTTG
TACTTTAAGAAGGCGAATTAGAACGAGACGCGGGAACAGGACTGGTGGTACCGAAAAAAC
M K F F R L I L L C A L V L T T M A>
SIGNAL PEPTIDE
M K F F R L I L L C A L V L T T M A F L>
CODING REGION
F L>
70 80 90 100 110 120
GCTCCAAGTACGGCAGCCAAGAGGCGTTGTGGCCGCCGCTTAATTCCCTATGTCTATTCA
CGAGGTTCATGCCGTCGGTTCTCCGCAACACCGGCGGCGAATTAAGGGATACAGATAAGT
A P S T A A K R R C G R R L I P Y V Y S>
CODING REGION
A P S T A A K R R C G R R L I P Y V Y S>
B DOMAIN
130 140 150 160 170 180
ATATGCGGCGGCCCGTGCGAGAATGGAGATATTATCATCGAGCACTGCTTCTCCGGAACA
TATACGCCGCCGGGCACGCTCTTACCTCTATAATAGTAGCTCGTGACGAAGAGGCCTTGT
I C G G P C E N G D I I I E H C F S G T>
CODING REGION
I C G G P C E N G D>
B DOMAIN
I I I E H C F S G T>
A DOMAIN
190 200 210 220 230 240
ACTCCCACCATTGCCGAAGTCCAAAAGGCTTGCTGTCCTGAACTATCTGAAGACCCAACT
TGAGGGTGGTAACGGCTTCAGGTTTTCCGAACGACAGGACTTGATAGACTTCTGGGTTGA
T P T I A E V Q K A C C P E L S E D P T>
CODING REGION
T P T I A E V Q K A C C P E L S E D P T>
A DOMAIN
250
TTCTCATCTTAA
AAGAGTAGAATT
F S S *>
F S S *>

FIG. 21

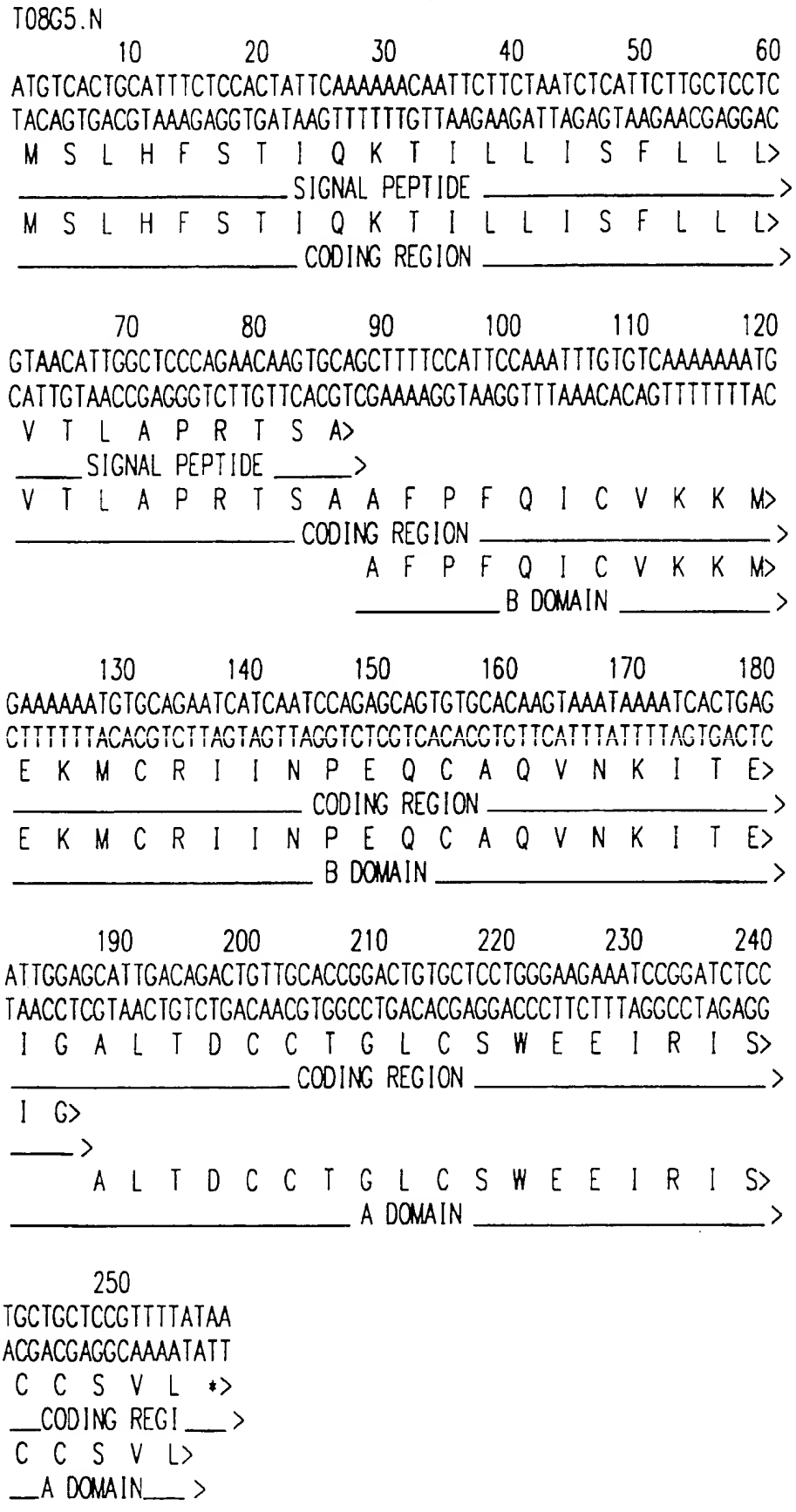

T08G5.N
10 20 30 40 50 60
ATGTCACTGCATTTCTCCACTATTCAAAAAACAATTCTTCTAATCTCATTCTTGCTCCTC
TACAGTGACGTAAAGAGGTGATAAGTTTTTTGTTAAGAAGATTAGAGTAAGAACGAGGAC
M S L H F S T I Q K T I L L I S F L L L>
SIGNAL PEPTIDE
M S L H F S T I Q K T I L L I S F L L L>
CODING REGION
70 80 90 100 110 120
GTAACATTGGCTCCCAGAACAAGTGCAGCTTTTCCATTCCAAATTTGTGTCAAAAAAATG
CATTGTAACCGAGGGTCTTGTTCACGTCGAAAAGGTAAGGTTTAAACACAGTTTTTTTAC
V T L A P R T S A>
SIGNAL PEPTIDE
V T L A P R T S A A F P F Q I C V K K M>
CODING REGION
A F P F Q I C V K K M>
B DOMAIN
130 140 150 160 170 180
GAAAAAATGTGCAGAATCATCAATCCAGAGCAGTGTGCACAAGTAAATAAAATCACTGAG
CTTTTTTACACGTCTTAGTAGTTAGGTCTCGTCACACGTCTTCATTTATTTTAGTGACTC
E K M C R I I N P E Q C A Q V N K I T E>
CODING REGION
E K M C R I I N P E Q C A Q V N K I T E>
B DOMAIN
190 200 210 220 230 240
ATTGGAGCATTGACAGACTGTTGCACCGGACTGTGCTCCTGGGAAGAAATCCGGATCTCC
TAACCTCGTAACTGTCTGACAACGTGGCCTGACACGAGGACCCTTCTTTAGGCCTAGAGG
I G A L T D C C T G L C S W E E I R I S>
CODING REGION
I G>
A L T D C C T G L C S W E E I R I S>
A DOMAIN
250
TGCTGCTCCGTTTTATAA
ACGACGAGGCAAAATATT
C C S V L *>
CODING REGI
C C S V L>
A DOMAIN

FIG. 22

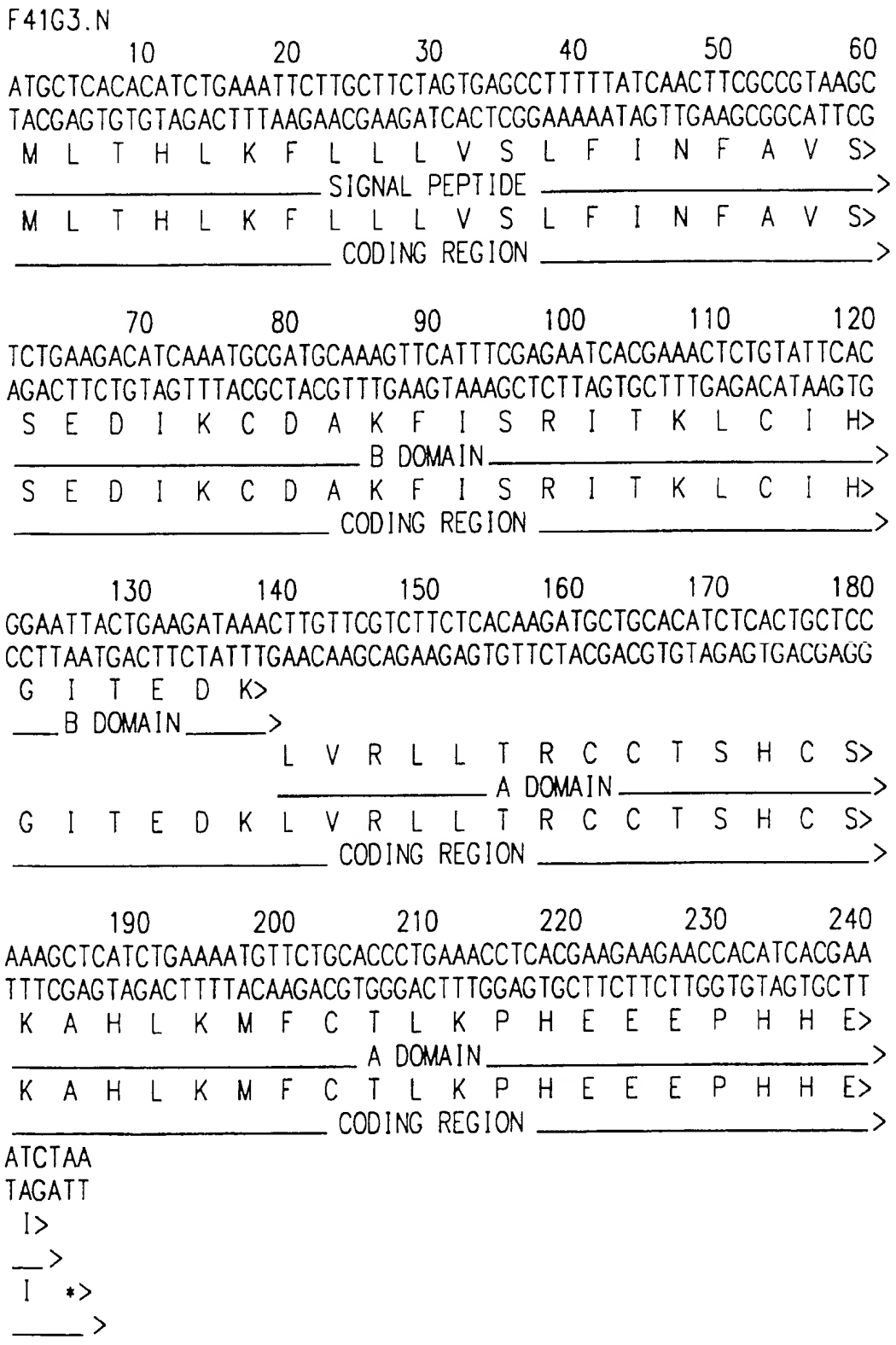

F41G3.N
10 20 30 40 50 60
ATGCTCACACATCTGAAATTCTTGCTTCTAGTGAGCCTTTTTATCAACTTCGCCGTAAGC
TACGAGTGTGTAGACTTTAAGAACGAAGATCACTCGGAAAAATAGTTGAAGCGGCATTCG
M L T H L K F L L L V S L F I N F A V S>
SIGNAL PEPTIDE
M L T H L K F L L L V S L F I N F A V S>
CODING REGION
70 80 90 100 110 120
TCTGAAGACATCAAATGCGATGCAAAGTTCATTTCGAGAATCACGAAACTCTGTATTCAC
AGACTTCTGTAGTTTACGCTACGTTTGAAGTAAAGCTCTTAGTGCTTTGAGACATAAGTG
S E D I K C D A K F I S R I T K L C I H>
B DOMAIN
S E D I K C D A K F I S R I T K L C I H>
CODING REGION
130 140 150 160 170 180
GGAATTACTGAAGATAAACTTGTTCGTCTTCTCACAAGATGCTGCACATCTCACTGCTCC
CCTTAATGACTTCTATTTGAACAAGCAGAAGAGTGTTCTACGACGTGTAGAGTGACGAGG
G I T E D K>
B DOMAIN
L V R L L T R C C T S H C S>
A DOMAIN
G I T E D K L V R L L T R C C T S H C S>
CODING REGION
190 200 210 220 230 240
AAAGCTCATCTGAAAATGTTCTGCACCCTGAAACCTCACGAAGAAGAACCACATCACGAA
TTTCGAGTAGACTTTTACAAGACGTGGGACTTTGGAGTGCTTCTTCTTGGTGTAGTGCTT
K A H L K M F C T L K P H E E E P H H E>
A DOMAIN
K A H L K M F C T L K P H E E E P H H E>
CODING REGION
ATCTAA
TAGATT
I>
I *>

FIG. 23

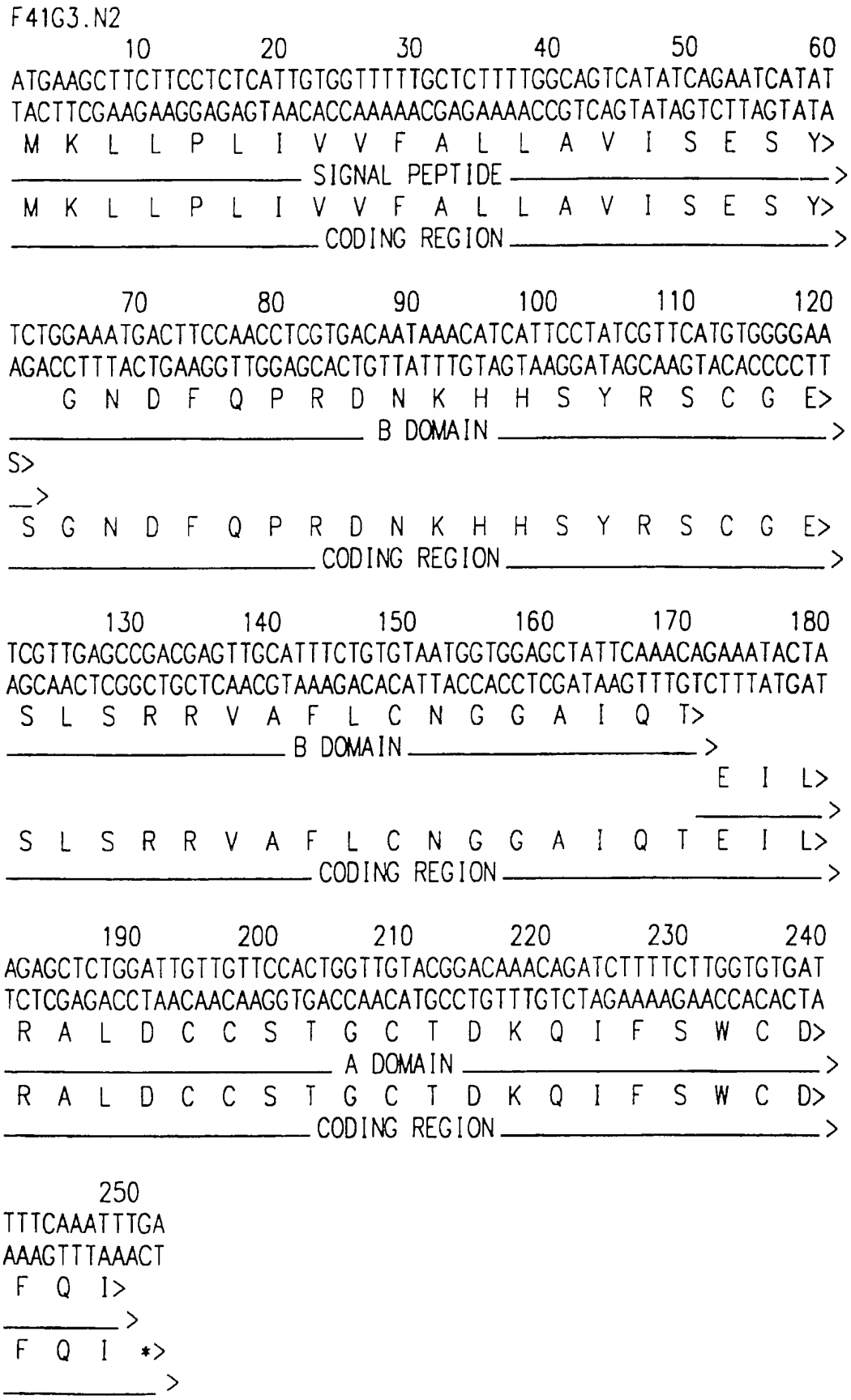

F41G3.N2
10 20 30 40 50 60
ATGAAGCTTCTTCCTCTCATTGTGGTTTTTGCTCTTTTGGCAGTCATATCAGAATCATAT
TACTTCGAAGAAGGAGAGTAACACCAAAAACGAGAAAACCGTCAGTATAGTCTTAGTATA
M K L L P L I V V F A L L A V I S E S Y>
SIGNAL PEPTIDE
M K L L P L I V V F A L L A V I S E S Y>
CODING REGION
70 80 90 100 110 120
TCTGGAAATGACTTCCAACCTCGTGACAATAAACATCATTCCTATCGTTCATGTGGGGAA
AGACCTTTACTGAAGGTTGGAGCACTGTTATTTGTAGTAAGGATAGCAAGTACACCCCTT
G N D F Q P R D N K H H S Y R S C G E>
B DOMAIN
S>
S G N D F Q P R D N K H H S Y R S C G E>
CODING REGION
130 140 150 160 170 180
TCGTTGAGCCGACGAGTTGCATTTCTGTGTAATGGTGGAGCTATTCAAACAGAAATACTA
AGCAACTCGGCTGCTCAACGTAAAGACACATTACCACCTCGATAAGTTTGTCTTTATGAT
S L S R R V A F L C N G G A I Q T>
B DOMAIN
E I L>
S L S R R V A F L C N G G A I Q T E I L>
CODING REGION
190 200 210 220 230 240
AGAGCTCTGGATTGTTGTTCCACTGGTTGTACGGACAAACAGATCTTTTCTTGGTGTGAT
TCTCGAGACCTAACAACAAGGTGACCAACATGCCTGTTTGTCTAGAAAAGAACCACACTA
R A L D C C S T G C T D K Q I F S W C D>
A DOMAIN
R A L D C C S T G C T D K Q I F S W C D>
CODING REGION
250
TTTCAAATTTGA
AAAGTTTAAACT
F Q I>
F Q I *>

FIG. 24

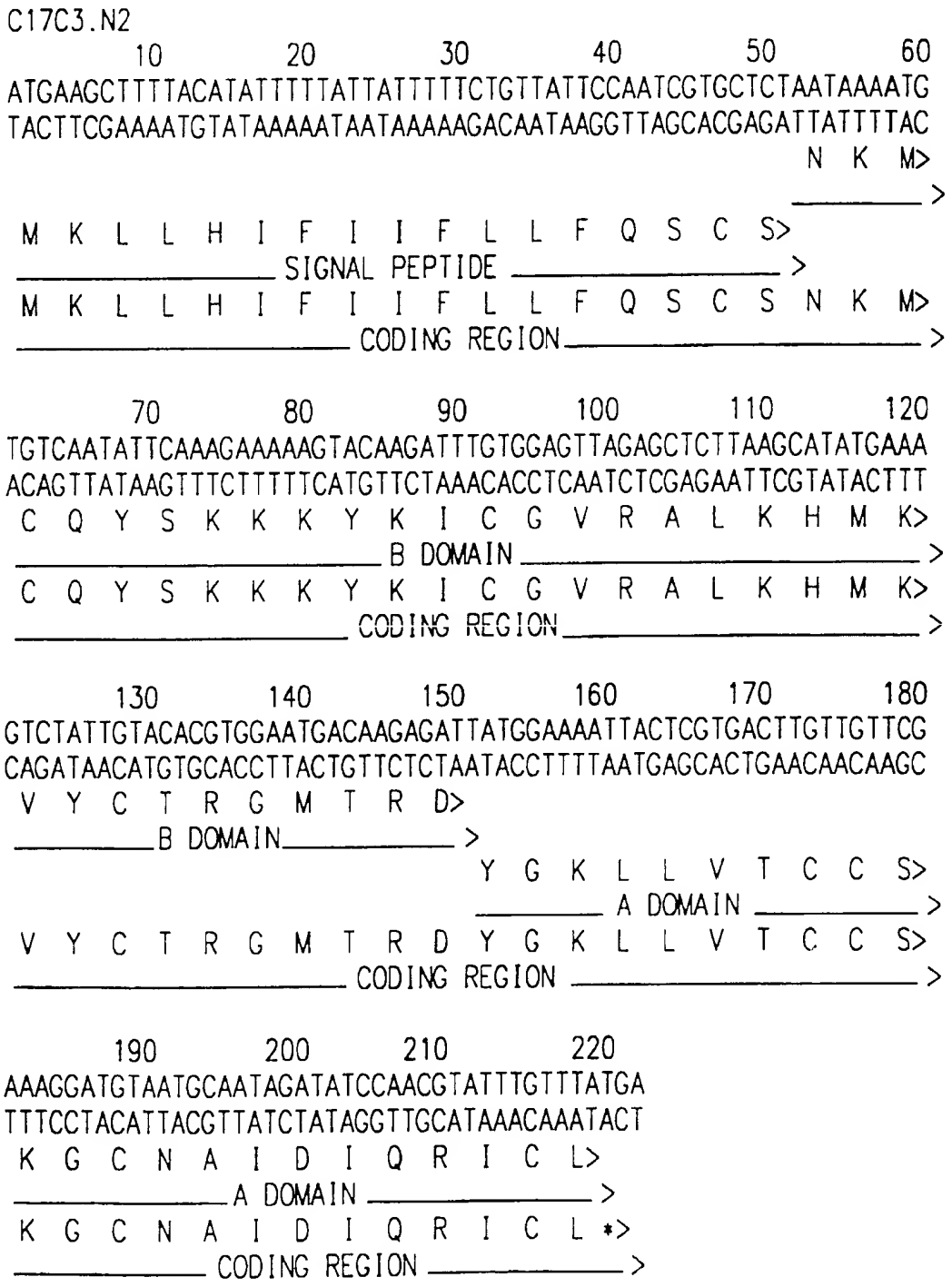

C17C3.N2
10 20 30 40 50 60
ATGAAGCTTTTACATATTTTTATTATTTTTCTGTTATTCCAATCGTGCTCTAATAAAATG
TACTTCGAAAATGTATAAAAATAATAAAAAGACAATAAGGTTAGCACGAGATTATTTTAC
N K M>
M K L L H I F I I F L L F Q S C S>
SIGNAL PEPTIDE
M K L L H I F I I F L L F Q S C S N K M>
CODING REGION
70 80 90 100 110 120
TGTCAATATTCAAAGAAAAAGTACAAGATTTGTGGAGTTAGAGCTCTTAAGCATATGAAA
ACAGTTATAAGTTTCTTTTTCATGTTCTAAACACCTCAATCTCGAGAATTCGTATACTTT
C Q Y S K K K Y K I C G V R A L K H M K>
B DOMAIN
C Q Y S K K K Y K I C G V R A L K H M K>
CODING REGION
130 140 150 160 170 180
GTCTATTGTACACGTGGAATGACAAGAGATTATGGAAAATTACTCGTGACTTGTTGTTCG
CAGATAACATGTGCACCTTACTGTTCTCTAATACCTTTTAATGAGCACTGAACAACAAGC
V Y C T R G M T R D>
B DOMAIN
Y G K L L V T C C S>
A DOMAIN
V Y C T R G M T R D Y G K L L V T C C S>
CODING REGION
190 200 210 220
AAAGGATGTAATGCAATAGATATCCAACGTATTTGTTTATGA
TTTCCTACATTACGTTATCTATAGGTTGCATAAACAAATACT
K G C N A I D I Q R I C L>
A DOMAIN
K G C N A I D I Q R I C L *>
CODING REGION

FIG. 25

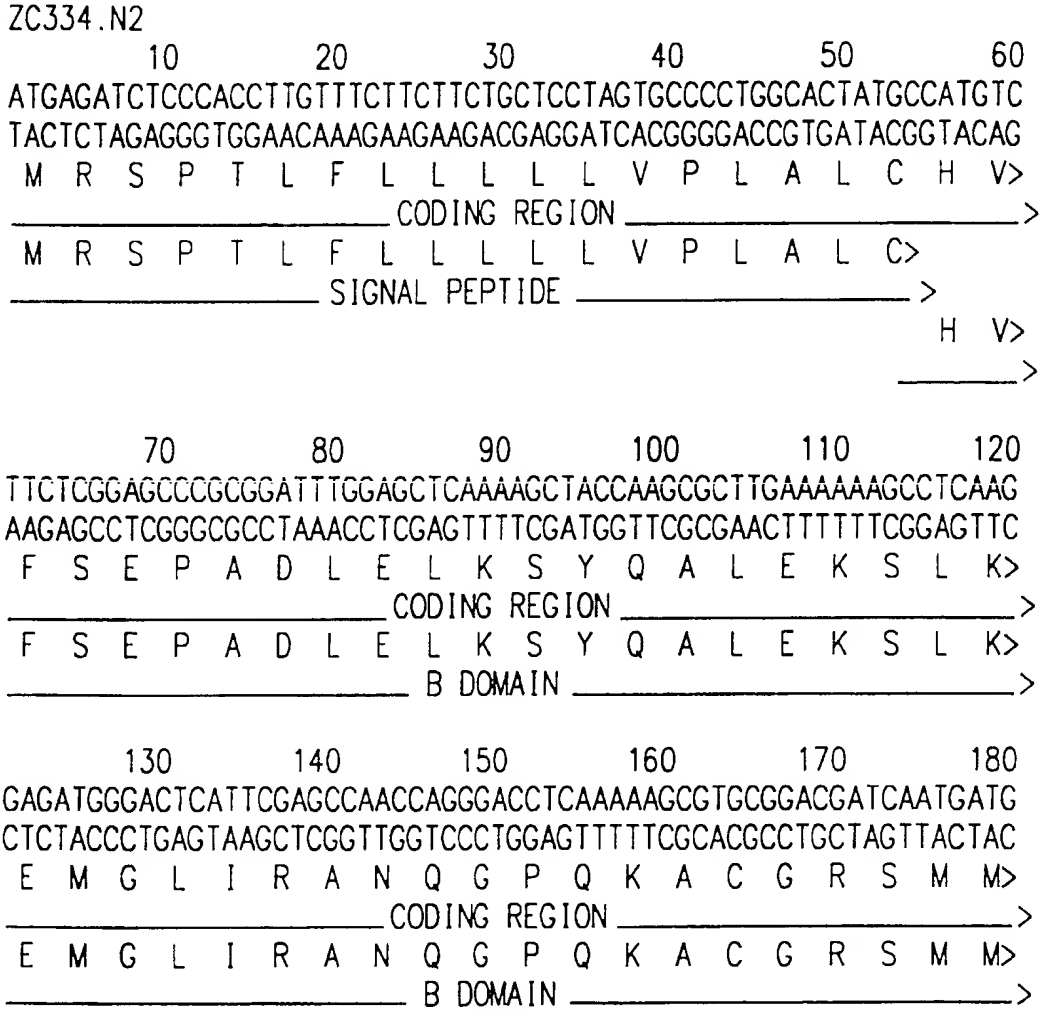

ZC334.N2
10 20 30 40 50 60
ATGAGATCTCCCACCTTGTTTCTTCTTCTGCTCCTAGTGCCCCTGGCACTATGCCATGTC
TACTCTAGAGGGTGGAACAAAGAAGAAGACGAGGATCACGGGGACCGTGATACGGTACAG
M R S P T L F L L L L L V P L A L C H V>
CODING REGION
M R S P T L F L L L L L V P L A L C>
SIGNAL PEPTIDE
H V>
70 80 90 100 110 120
TTCTCGGAGCCCGCGGATTTGGAGCTCAAAAGCTACCAAGCGCTTGAAAAAAAGCCTCAAG
AAGAGCCTCGGGCGCCTAAACCTCGAGTTTTCGATGGTTCGCGAACTTTTTTCGGAGTTC
F S E P A D L E L K S Y Q A L E K S L K>
CODING REGION
F S E P A D L E L K S Y Q A L E K S L K>
B DOMAIN
130 140 150 160 170 180
GAGATGGGACTCATTCGAGCCAACCAGGGACCTCAAAAAGCGTGCGGACGATCAATGATG
CTCTACCCTGAGTAAGCTCGGTTGGTCCCTGGAGTTTTTCGCACGCCTGCTAGTTACTAC
E M G L I R A N Q G P Q K A C G R S M M>
CODING REGION
E M G L I R A N Q G P Q K A C G R S M M>
B DOMAIN

FIG. 26A

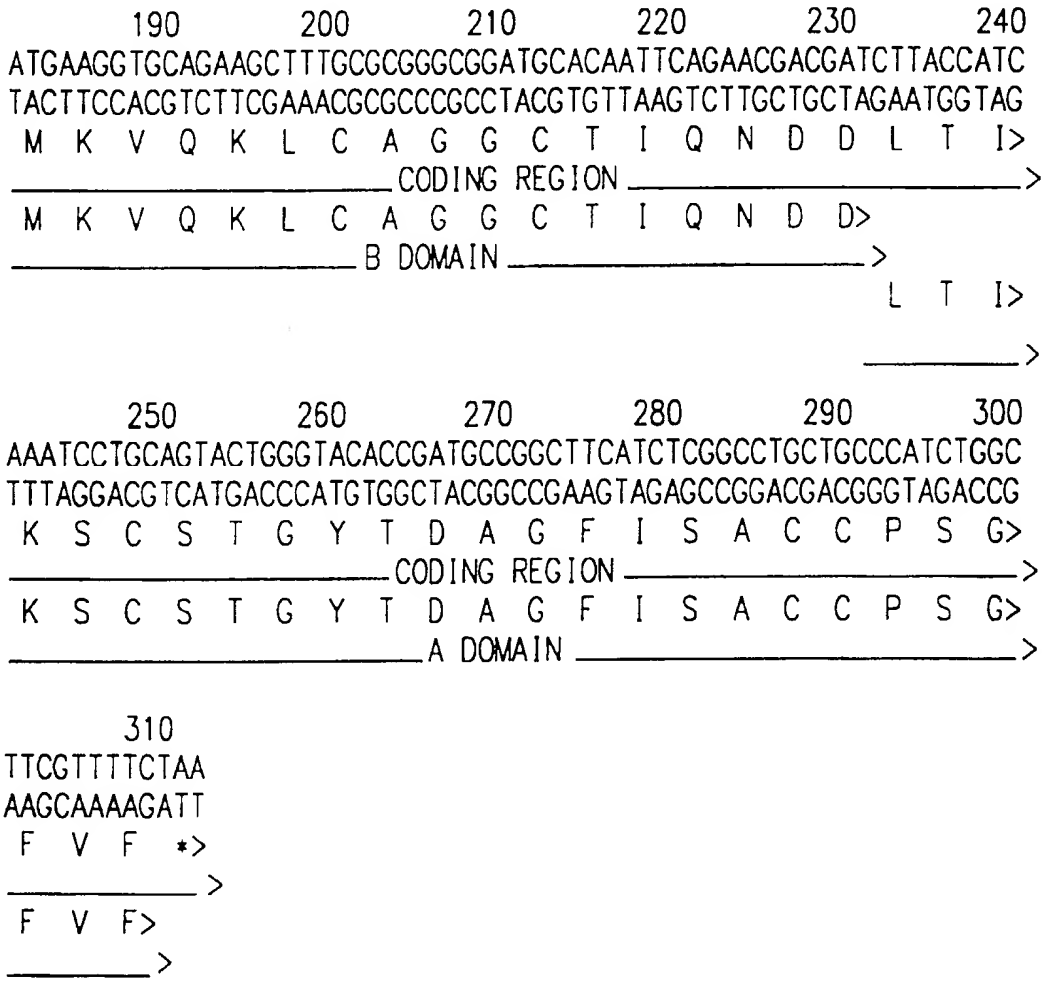

190 200 210 220 230 240
ATGAAGGTGCAGAAGCTTTGCGCGGGCGGATGCACAATTCAGAACGACGATCTTACCATC
TACTTCCACGTCTTCGAAACGCGCCCGCCTACGTGTTAAGTCTTGCTGCTAGAATGGTAG
M K V Q K L C A G G C T I Q N D D L T I>
CODING REGION
M K V Q K L C A G G C T I Q N D D>
B DOMAIN
L T I>
250 260 270 280 290 300
AAATCCTGCAGTACTGGGTACACCGATGCCGGCTTCATCTCGGCCTGCTGCCCATCTGGC
TTTAGGACGTCATGACCCATGTGGCTACGGCCGAAGTAGAGCCGGACGACGGGTAGACCG
K S C S T G Y T D A G F I S A C C P S G>
CODING REGION
K S C S T G Y T D A G F I S A C C P S G>
A DOMAIN
310
TTCGTTTTCTAA
AAGCAAAAGATT
F V F *>
F V F>
F V F>
F V F>

FIG. 26B

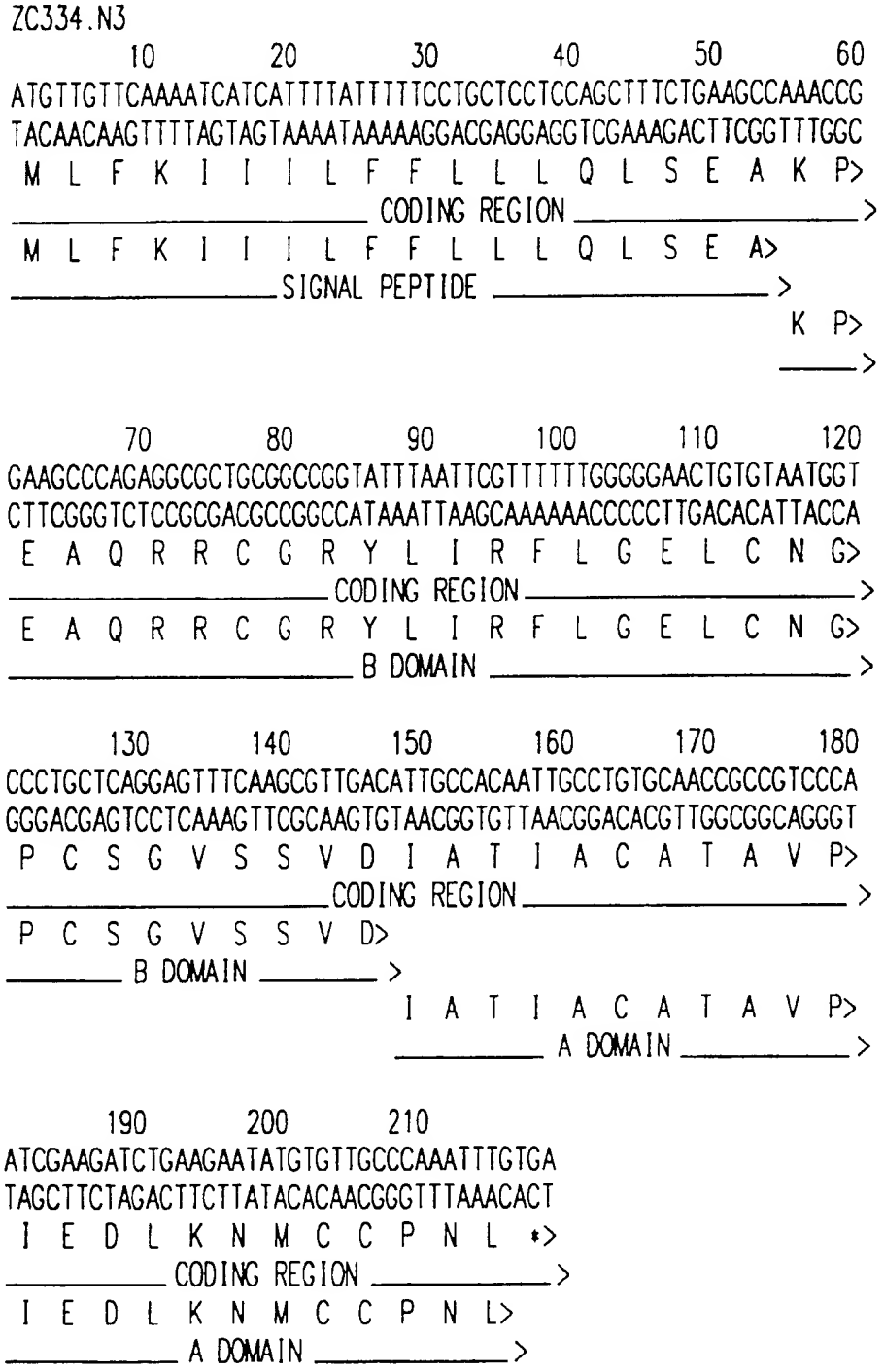

ZC334.N3
10 20 30 40 50 60
ATGTTGTTCAAAATCATCATTTTATTTTTCCTGCTCCTCCAGCTTTCTGAAGCCAAACCG
TACAACAAGTTTTAGTAGTAAAATAAAAAGGACGAGGAGGTCGAAAGACTTCGGTTTGGC
M L F K I I I L F F L L L Q L S E A K P>
CODING REGION
M L F K I I I L F F L L L Q L S E A>
SIGNAL PEPTIDE
K P>
70 80 90 100 110 120
GAAGCCCAGAGGCGCTGCGGCCCGTATTTAATTCGTTTTTTGGGGGAACTGTGTAATGGT
CTTCGGGTCTCCGCGACGCCGGCCATAAATTAAGCAAAAAACCCCCTTGACACATTACCA
E A Q R R C G R Y L I R F L G E L C N G>
CODING REGION
E A Q R R C G R Y L I R F L G E L C N G>
B DOMAIN
130 140 150 160 170 180
CCCTGCTCAGGAGTTTCAAGCGTTGACATTGCCACAATTGCCTGTGCAACCGCCGTCCCA
GGGACGAGTCCTCAAAGTTCGCAAGTGTAACGGTGTTAACGGACACGTTGGCGGCAGGGT
P C S G V S S V D I A T I A C A T A V P>
CODING REGION
P C S G V S S V D>
B DOMAIN
I A T I A C A T A V P>
A DOMAIN
190 200 210
ATCGAAGATCTGAAGAATATGTGTTGCCCAAATTTGTGA
TAGCTTCTAGACTTCTTATACACAACGGGTTTAAACACT
I E D L K N M C C P N L *>
CODING REGION
I E D L K N M C C P N L>
A DOMAIN

FIG. 27

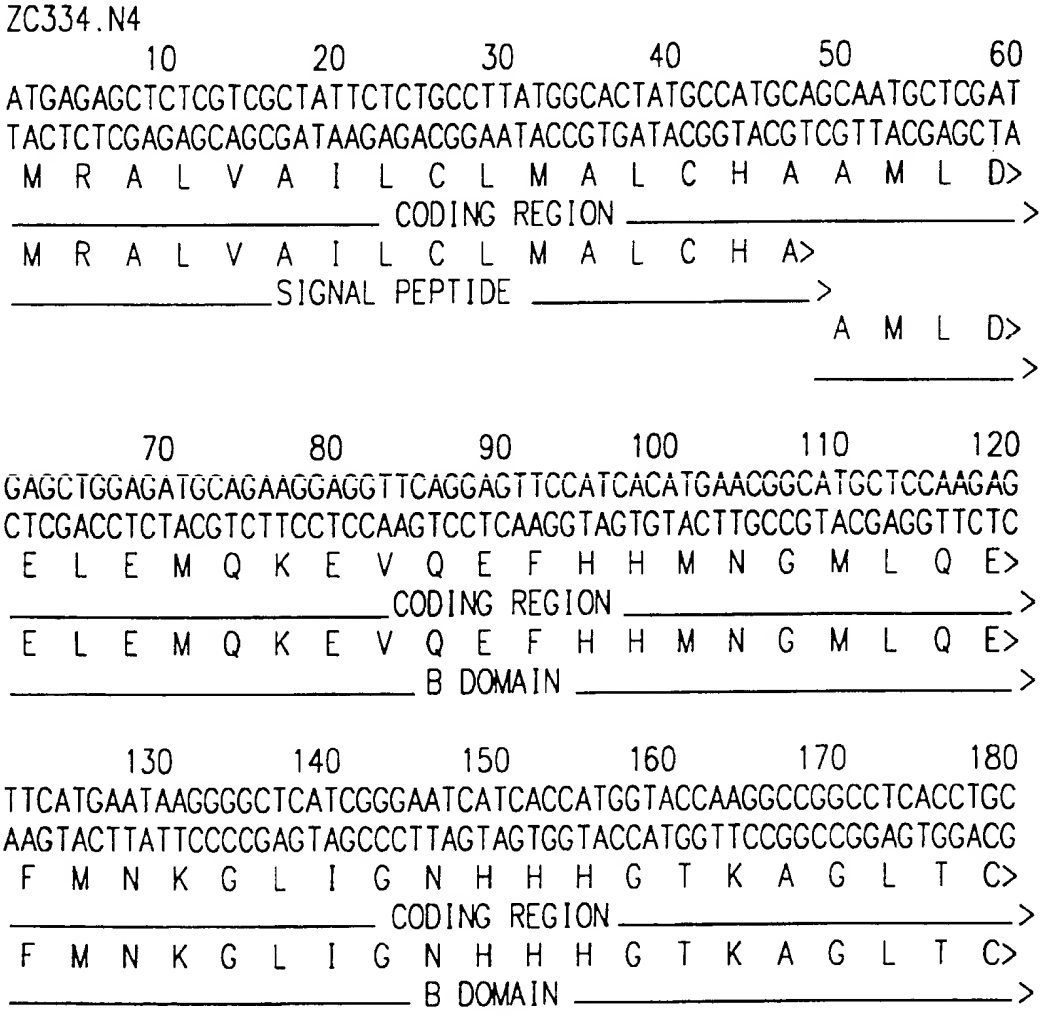

ZC334.N4
10 20 30 40 50 60
ATGAGAGCTCTCGTCGCTATTCTCTGCCTTATGGCACTATGCCATGCAGCAATGCTCGAT
TACTCTCGAGAGCAGCGATAAGAGACGGAATACCGTGATACGGTACGTCGTTACGAGCTA
M R A L V A I L C L M A L C H A A M L D>
CODING REGION
M R A L V A I L C L M A L C H A>
SIGNAL PEPTIDE
A M L D>
70 80 90 100 110 120
GAGCTGGAGATGCAGAAGGAGGTTCAGGAGTTCCATCACATGAACGGCATGCTCCAAGAG
CTCGACCTCTACGTCTTCCTCCAAGTCCTCAAGGTAGTGTACTTGCCGTACGAGGTTCTC
E L E M Q K E V Q E F H H M N G M L Q E>
CODING REGION
E L E M Q K E V Q E F H H M N G M L Q E>
B DOMAIN
130 140 150 160 170 180
TTCATGAATAAGGGGCTCATCGGGAATCATCACCATGGTACCAAGGCCGGCCTCACCTGC
AAGTACTTATTCCCCGAGTAGCCCTTAGTAGTGGTACCATGGTTCCGGCCGGAGTGGACG
F M N K G L I G N H H H G T K A G L T C>
CODING REGION
F M N K G L I G N H H H G T K A G L T C>
B DOMAIN

FIG. 28A

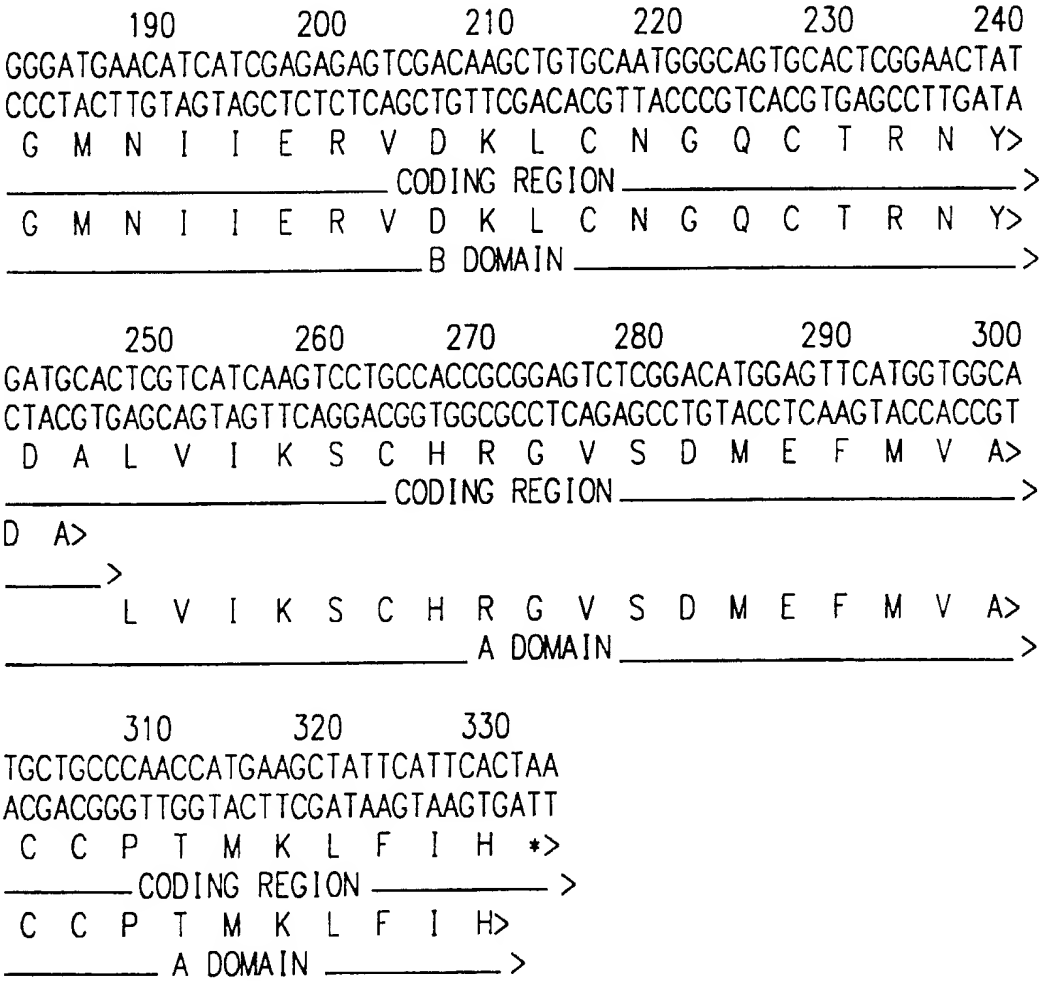

190 200 210 220 230 240
GGGATGAACATCATCGAGAGAGTCGACAAGCTGTGCAATGGGCAGTGCACTCGGAACTAT
CCCTACTTGTAGTAGCTCTCTCAGCTGTTCGACACGTTACCCGTCACGTGAGCCTTGATA
G M N I I E R V D K L C N G Q C T R N Y>
CODING REGION
G M N I I E R V D K L C N G Q C T R N Y>
B DOMAIN
250 260 270 280 290 300
GATGCACTCGTCATCAAGTCCTGCCACCGCGGAGTCTCGGACATGGAGTTCATGGTGGCA
CTACGTGAGCAGTAGTTCAGGACGGTGGCGCCTCAGAGCCTGTACCTCAAGTACCACCGT
D A L V I K S C H R G V S D M E F M V A>
CODING REGION
D A>
L V I K S C H R G V S D M E F M V A>
A DOMAIN
310 320 330
TGCTGCCCAACCATGAAGCTATTCATTCACTAA
ACGACGGGTTGGTACTTCGATAAGTAAGTGATT
C C P T M K L F I H *>
CODING REGION
C C P T M K L F I H>
A DOMAIN

FIG. 28B

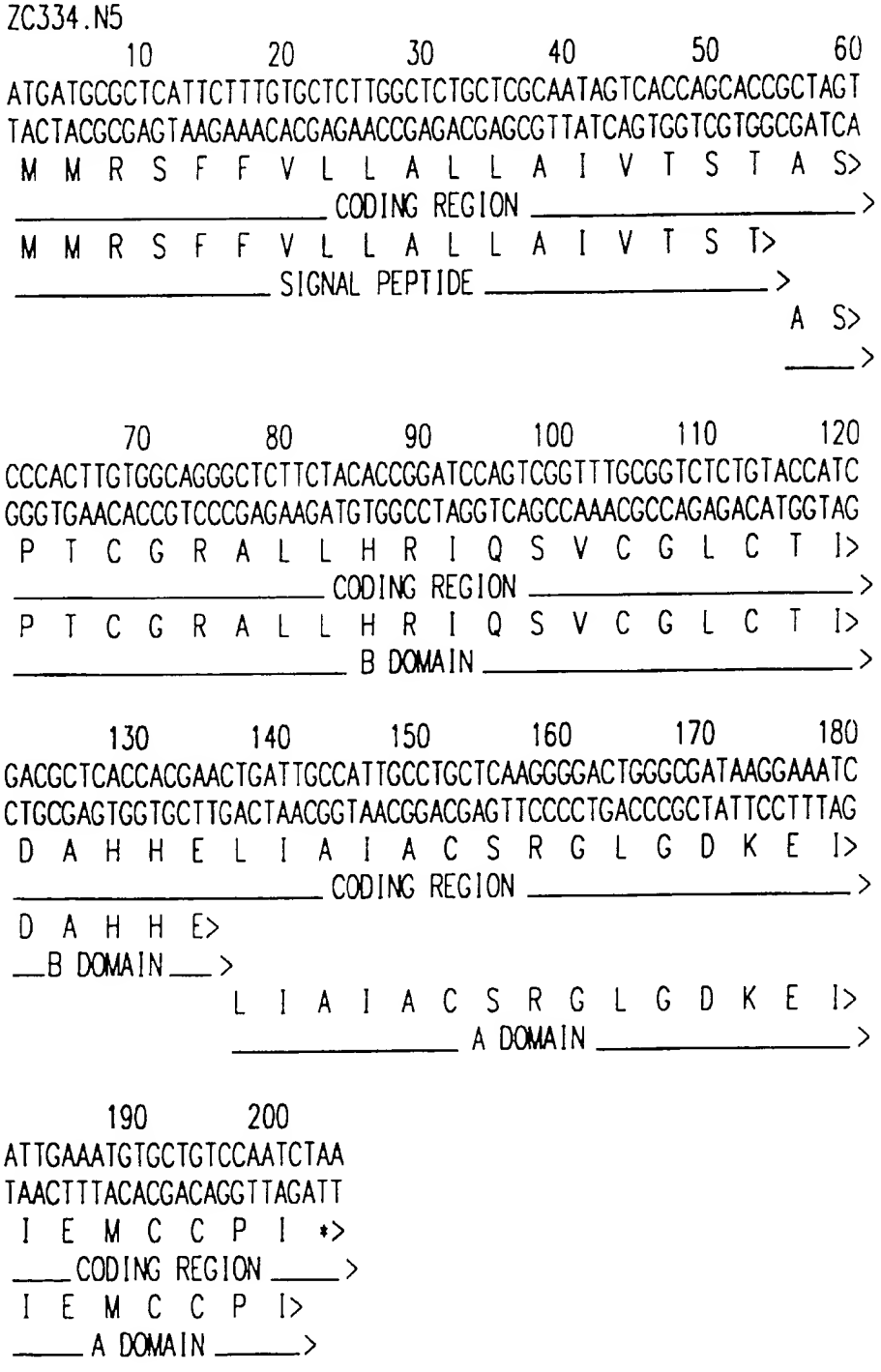

ZC334.N5
10 20 30 40 50 60
ATGATGCGCTCATTCTTTGTGCTCTTGGCTCTGCTCGCAATAGTCACCAGCACCGCTAGT
TACTACGCGAGTAAGAAACACGAGAACCGAGACGAGCGTTATCAGTGGTCGTGGCGATCA
M M R S F F V L L A L L A I V T S T A S>
CODING REGION
M M R S F F V L L A L L A I V T S T>
SIGNAL PEPTIDE
A S>
70 80 90 100 110 120
CCCACTTGTGGCAGGGCTCTTCTACACCGGATCCAGTCGGTTTGCGGTCTCTGTACCATC
GGGTGAACACCGTCCCGAGAAGATGTGGCCTAGGTCAGCCAAACGCCAGAGACATGGTAG
P T C G R A L L H R I Q S V C G L C T I>
CODING REGION
P T C G R A L L H R I Q S V C G L C T I>
B DOMAIN
130 140 150 160 170 180
GACGCTCACCACGAACTGATTGCCATTGCCTGCTCAAGGGGACTGGGCGATAAGGAAATC
CTGCGAGTGGTGCTTGACTAACGGTAACGGACGAGTTCCCCTGACCCGCTATTCCTTTAG
D A H H E L I A I A C S R G L G D K E I>
CODING REGION
D A H H E>
B DOMAIN
L I A I A C S R G L G D K E I>
A DOMAIN
190 200
ATTGAAATGTGCTGTCCAATCTAA
TAACTTTACACGACAGGTTAGATT
I E M C C P I *>
CODING REGION
I E M C C P I>
A DOMAIN

FIG. 29

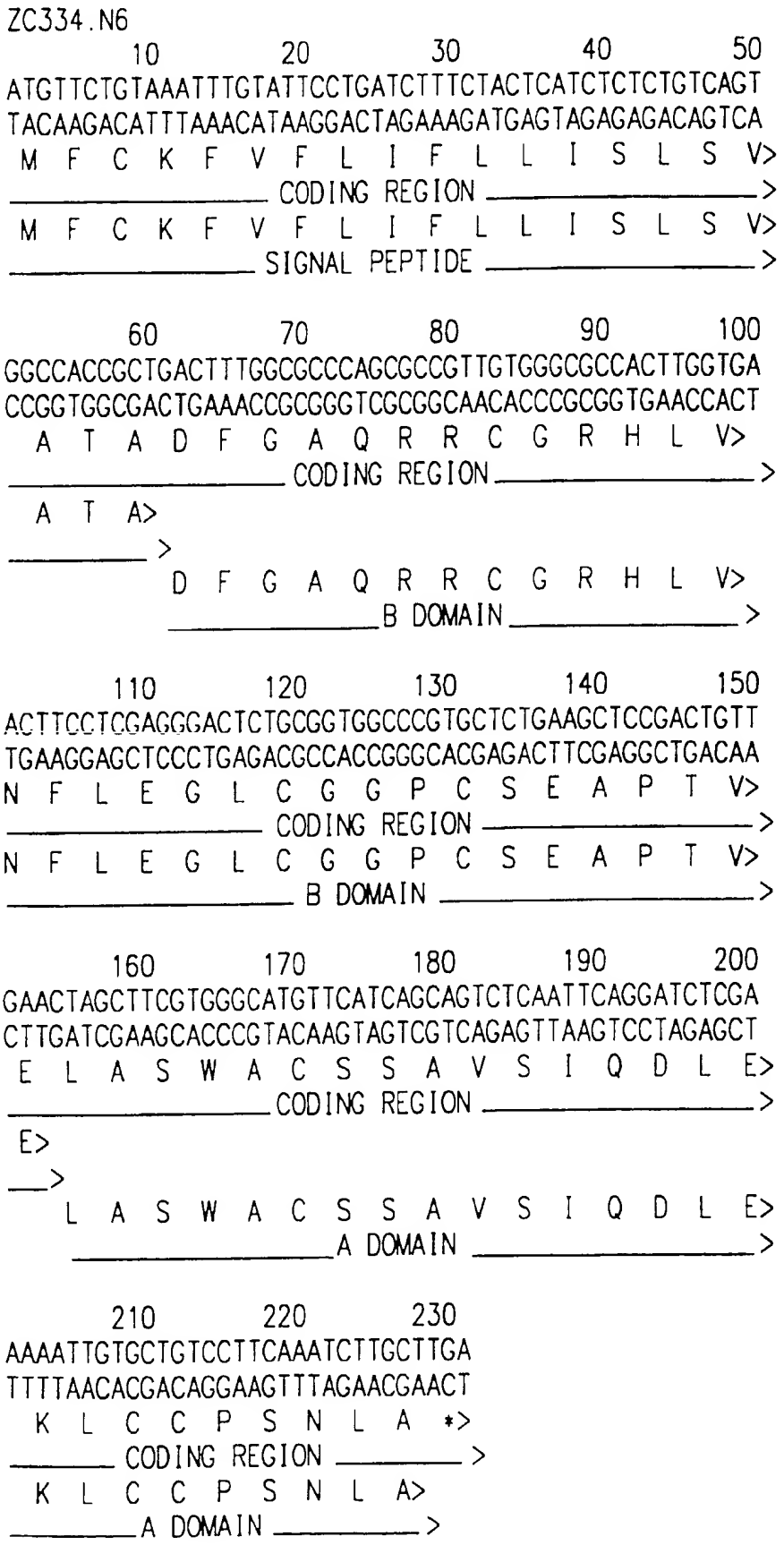

ZC334.N6
10 20 30 40 50
ATGTTCTGTAAATTTGTATTCCTGATCTTTCTACTCATCTCTCTGTCAGT
TACAAGACATTTAAACATAAGGACTAGAAAGATGAGTAGAGAGACAGTCA
M F C K F V F L I F L L I S L S V>
CODING REGION
M F C K F V F L I F L L I S L S V>
SIGNAL PEPTIDE
60 70 80 90 100
GGCCACCGCTGACTTTGGCGCCCAGCGCCGTTGTGGGCGCCACTTGGTGA
CCGGTGGCGACTGAAACCGCGGGTCGCGGCAACACCCGCGGTGAACCACT
A T A D F G A Q R R C G R H L V>
CODING REGION
A T A>
D F G A Q R R C G R H L V>
B DOMAIN
110 120 130 140 150
ACTTCCTCGAGGGACTCTGCGGTGGCCCGTGCTCTGAAGCTCCGACTGTT
TGAAGGAGCTCCCTGAGACGCCACCGGGCACGAGACTTCGAGGCTGACAA
N F L E G L C G G P C S E A P T V>
CODING REGION
N F L E G L C G G P C S E A P T V>
B DOMAIN
160 170 180 190 200
GAACTAGCTTCGTGGGCATGTTCATCAGCAGTCTCAATTCAGGATCTCGA
CTTGATCGAAGCACCCGTACAAGTAGTCGTCAGAGTTAAGTCCTAGAGCT
E L A S W A C S S A V S I Q D L E>
CODING REGION
E>
L A S W A C S S A V S I Q D L E>
A DOMAIN
210 220 230
AAAATTGTGCTGTCCTTCAAATCTTGCTTGA
TTTTAACACGACAGGAAGTTTAGAACGAACT
K L C C P S N L A *>
CODING REGION
K L C C P S N L A>
A DOMAIN

FIG. 30

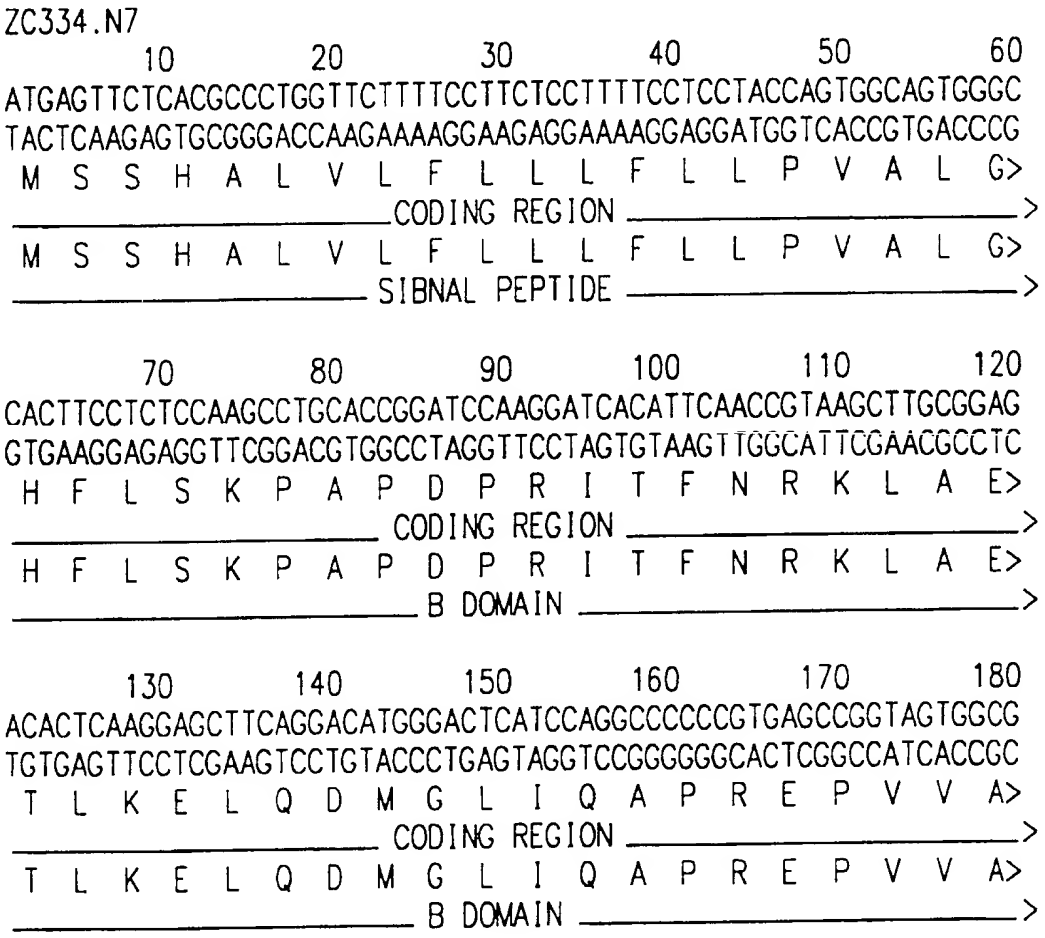

ZC334.N7
10 20 30 40 50 60
ATGAGTTCTCACGCCCTGGTTCTTTTCCTTCTCCTTTTCCTCCTACCAGTGGCAGTGGGC
TACTCAAGAGTGCGGGACCAAGAAAAGGAAGAGGAAAAGGAGGATGGTCACCGTGACCCG
M S S H A L V L F L L L F L L P V A L G>
CODING REGION
M S S H A L V L F L L L F L L P V A L G>
SIBNAL PEPTIDE
70 80 90 100 110 120
CACTTCCTCTCCAAGCCTGCACCGGATCCAAGGATCACATTCAACCGTAAGCTTGCGGAG
GTGAAGGAGAGGTTCGGACGTGGCCTAGGTTCCTAGTGTAAGTTGGCATTCGAACGCCTC
H F L S K P A P D P R I T F N R K L A E>
CODING REGION
H F L S K P A P D P R I T F N R K L A E>
B DOMAIN
130 140 150 160 170 180
ACACTCAAGGAGCTTCAGGACATGGGACTCATCCAGGCCCCCCCGTGAGCCGGTAGTGGCG
TGTGAGTTCCTCGAAGTCCTGTACCCTGAGTAGGTCCGGGGGGCACTCGGCCATCACCGC
T L K E L Q D M G L I Q A P R E P V V A>
CODING REGION
T L K E L Q D M G L I Q A P R E P V V A>
B DOMAIN

FIG. 31A

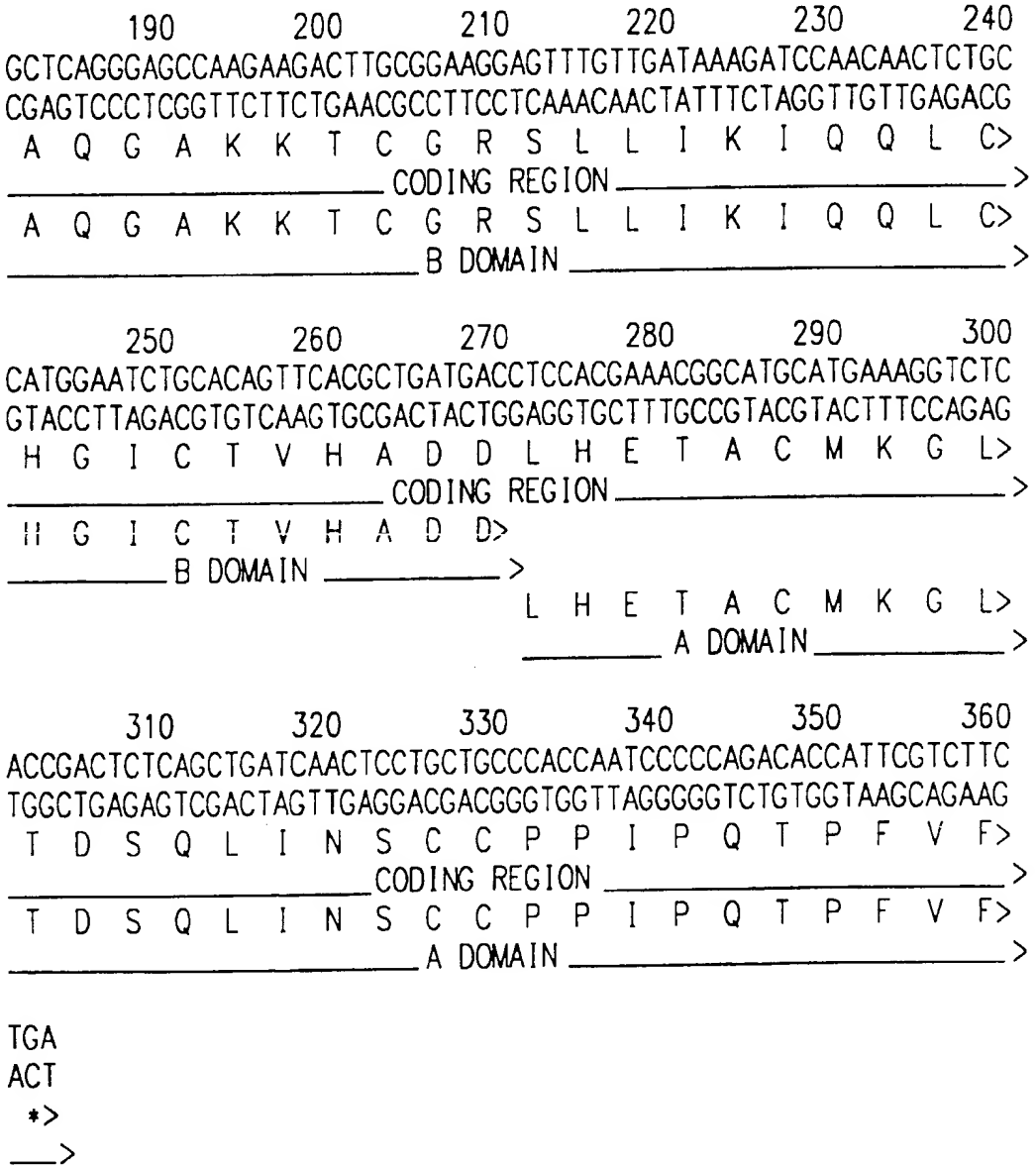

190 200 210 220 230 240
GCTCAGGGAGCCAAGAAGACTTGCGGAAGGAGTTTGTTGATAAAGATCCAACAACTCTGC
CGAGTCCCTCGGTTCTTCTGAACGCCTTCCTCAAACAACTATTTCTAGGTTGTTGAGACG
A Q G A K K T C G R S L L I K I Q Q L C>
CODING REGION
A Q G A K K T C G R S L L I K I Q Q L C>
B DOMAIN
250 260 270 280 290 300
CATGGAATCTGCACAGTTCACGCTGATGACCTCCACGAAACGGCATGCATGAAAGGTCTC
GTACCTTAGACGTGTCAAGTGCGACTACTGGAGGTGCTTTGCCGTACGTACTTTCCAGAG
H G I C T V H A D D L H E T A C M K G L>
CODING REGION
H G I C T V H A D D>
B DOMAIN
L H E T A C M K G L>
A DOMAIN
310 320 330 340 350 360
ACCGACTCTCAGCTGATCAACTCCTGCTGCCCACCAATCCCCCAGACACCATTCGTCTTC
TGGCTGAGAGTCGACTAGTTGAGGACGACGGGTGGTTAGGGGGTCTGTGGTAAGCAGAAG
T D S Q L I N S C C P P I P Q T P F V F>
CODING REGION
T D S Q L I N S C C P P I P Q T P F V F>
A DOMAIN
TGA
ACT
*>

FIG. 31B

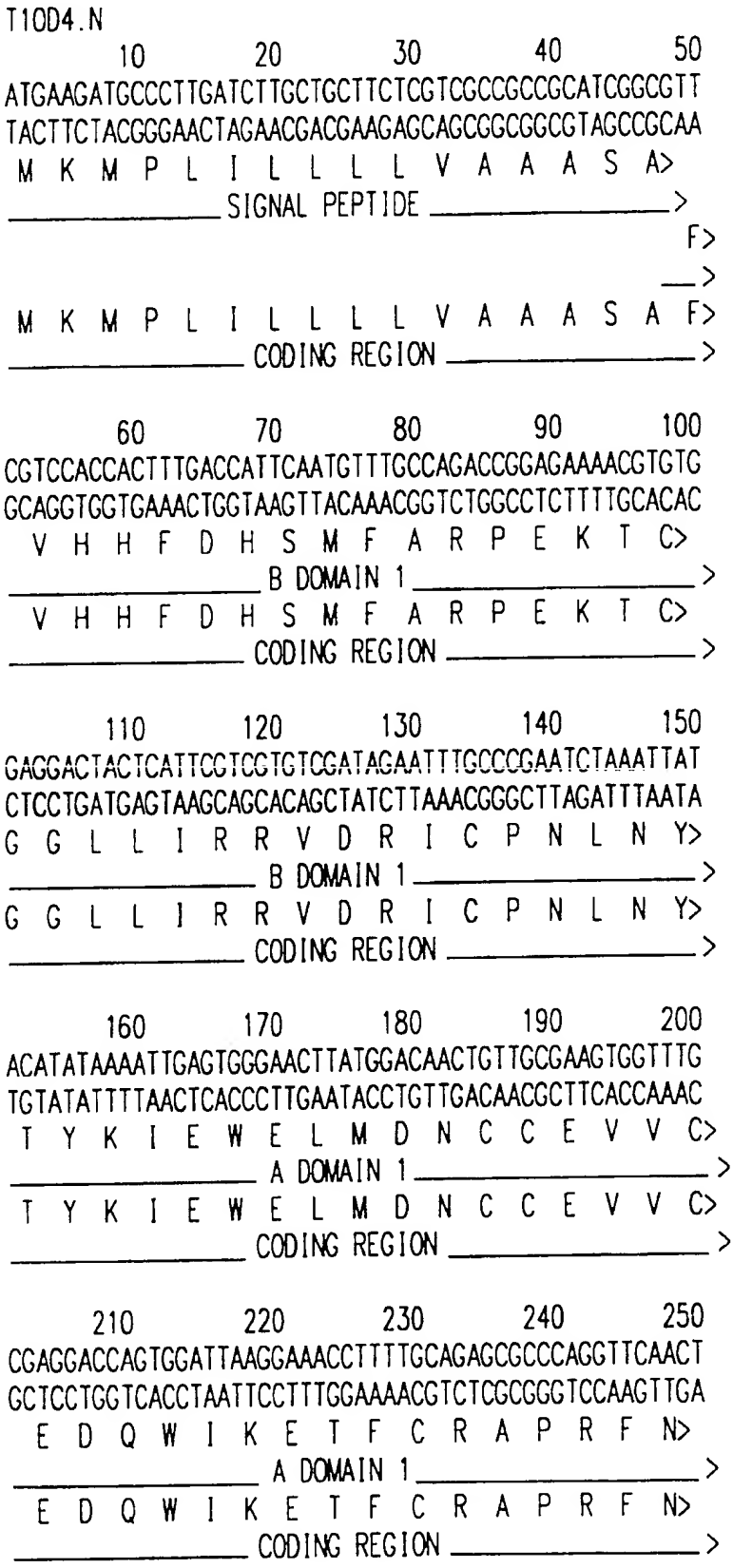

T10D4.N
10 20 30 40 50
ATGAAGATGCCCTTGATCTTGCTGCTTCTCGTCGCCGCCGCATCGGCGTT
TACTTCTACGGGAACTAGAACGACGAAGAGCAGCGGCGGCGTAGCCGCAA
M K M P L I L L L L V A A A S A>
SIGNAL PEPTIDE
F>
M K M P L I L L L L V A A A S A F>
CODING REGION
60 70 80 90 100
CGTCCACCACTTTGACCATTCAATGTTTGCCAGACCGGAGAAAACGTGTG
GCAGGTGGTGAAACTGGTAAGTTACAAACGGTCTGGCCTCTTTTGCACAC
V H H F D H S M F A R P E K T C>
B DOMAIN 1
V H H F D H S M F A R P E K T C>
CODING REGION
110 120 130 140 150
GAGGACTACTCATTCCTCCTGTCCATAGAATTTGCCCCGAATCTAAATTAT
CTCCTGATGAGTAAGCAGCACAGCTATCTTAAACGGGCTTAGATTTAATA
G G L L I R R V D R I C P N L N Y>
B DOMAIN 1
G G L L I R R V D R I C P N L N Y>
CODING REGION
160 170 180 190 200
ACATATAAAATTGAGTGGGAACTTATGGACAACTGTTGCGAAGTGGTTTG
TGTATATTTTAACTCACCCTTGAATACCTGTTGACAACGCTTCACCAAAC
T Y K I E W E L M D N C C E V V C>
A DOMAIN 1
T Y K I E W E L M D N C C E V V C>
CODING REGION
210 220 230 240 250
CGAGGACCAGTGGATTAAGGAAACCTTTTGCAGAGCGCCCAGGTTCAACT
GCTCCTGGTCACCTAATTCCTTTGGAAAACGTCTCGCGGGTCCAAGTTGA
E D Q W I K E T F C R A P R F N>
A DOMAIN 1
E D Q W I K E T F C R A P R F N>
CODING REGION

FIG. 32A

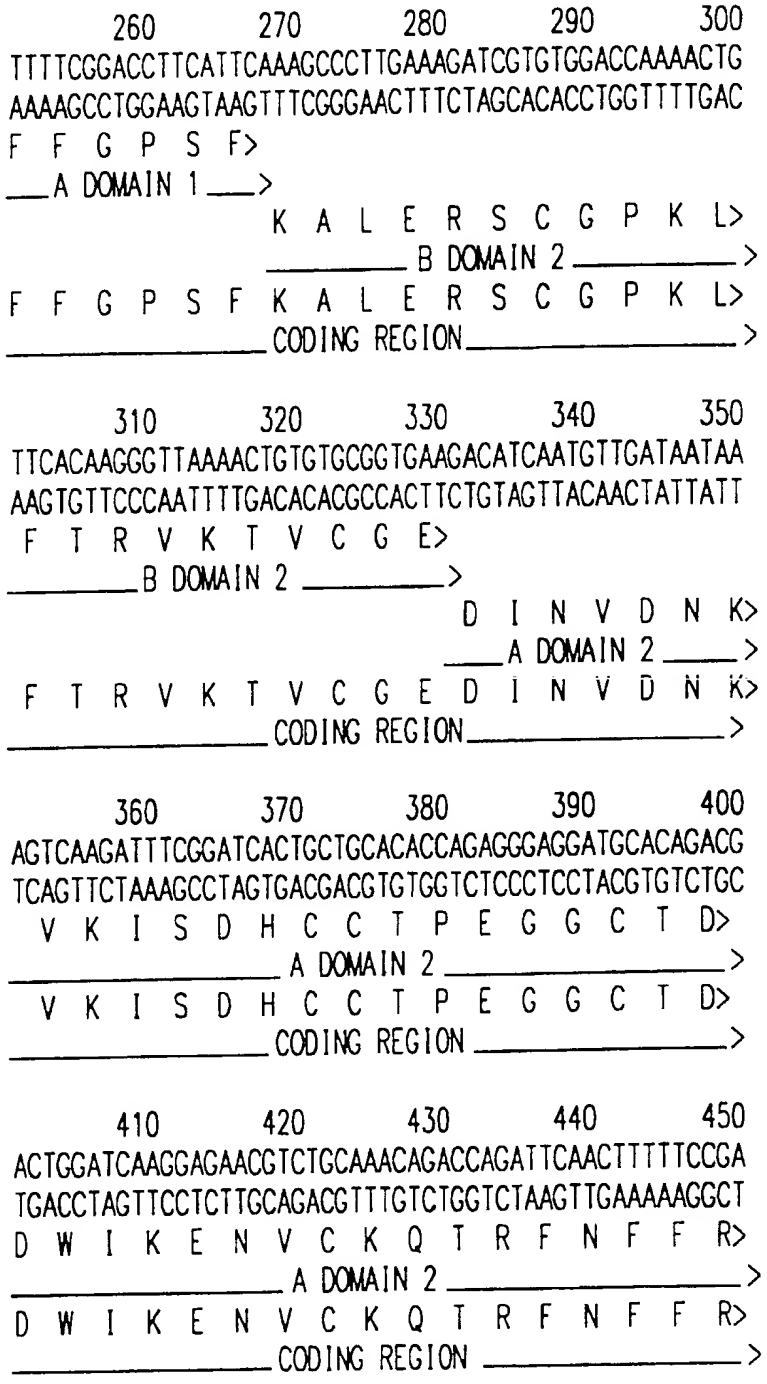

260 270 280 290 300
TTTTCGGACCTTCATTCAAAGCCCTTGAAAGATCGTGTGGACCAAAACTG
AAAAGCCTGGAAGTAAGTTTCGGGAACTTTCTAGCACACCTGGTTTTGAC
F F G P S F>
A DOMAIN 1 >
K A L E R S C G P K L>
B DOMAIN 2 >
F F G P S F K A L E R S C G P K L>
CODING REGION >
310 320 330 340 350
TTCACAAGGGTTAAAACTGTGTGCGGTGAAGACATCAATGTTGATAATAA
AAGTGTTCCCAATTTTGACACACGCCACTTCTGTAGTTACAACTATTATT
F T R V K T V C G E>
B DOMAIN 2 >
D I N V D N K>
A DOMAIN 2 >
F T R V K T V C G E D I N V D N K>
CODING REGION >
360 370 380 390 400
AGTCAAGATTTCGGATCACTGCTGCACACCAGAGGGAGGATGCACAGACG
TCAGTTCTAAAGCCTAGTGACGACGTGTGGTCTCCCTCCTACGTGTCTGC
V K I S D H C C T P E G G C T D>
A DOMAIN 2 >
V K I S D H C C T P E G G C T D>
CODING REGION >
410 420 430 440 450
ACTGGATCAAGGAGAACGTCTGCAAACAGACCAGATTCAACTTTTTCCGA
TGACCTAGTTCCTCTTGCAGACGTTTGTCTGGTCTAAGTTGAAAAAGGCT
D W I K E N V C K Q T R F N F F R>
A DOMAIN 2 >
D W I K E N V C K Q T R F N F F R>
CODING REGION >

FIG. 32B

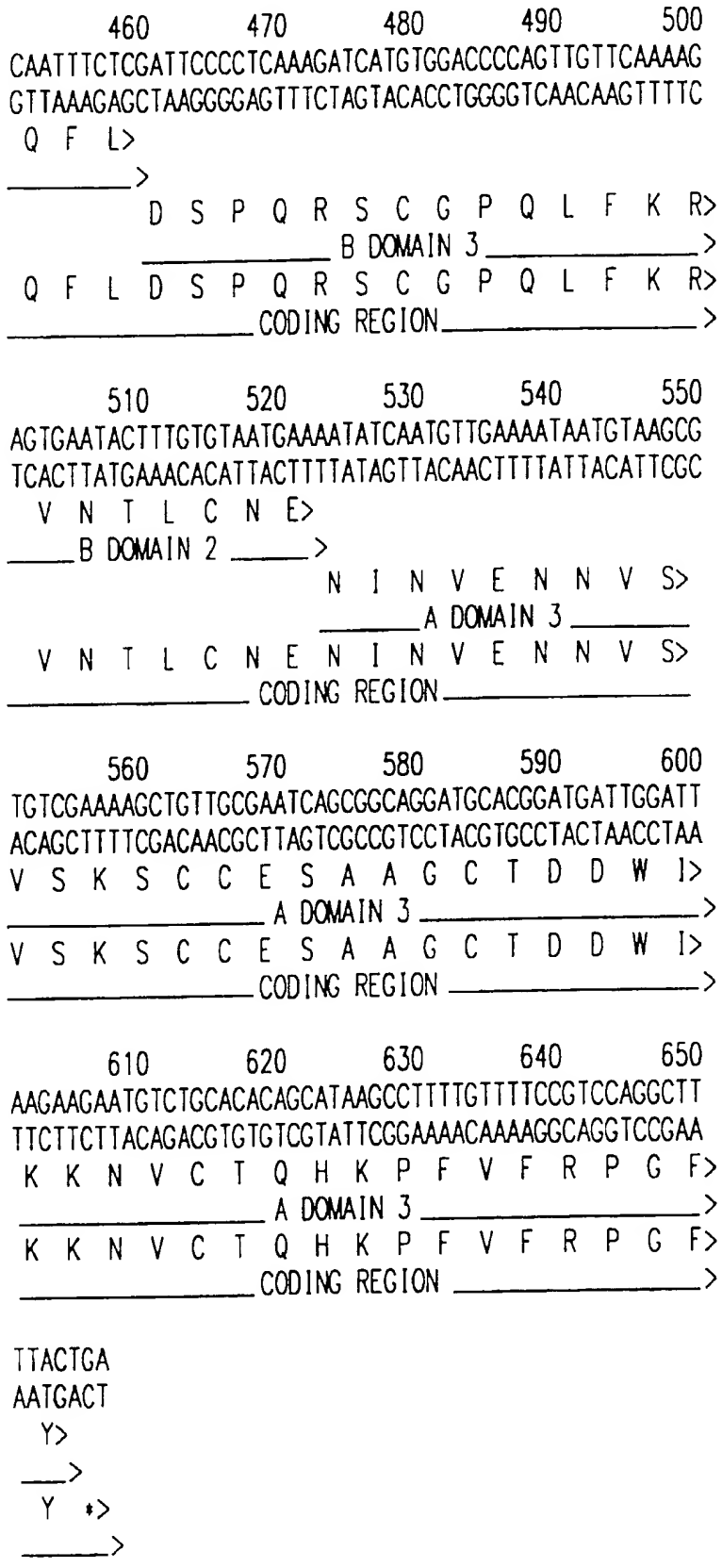

460 470 480 490 500
CAATTTCTCGATTCCCCTCAAAGATCATGTGGACCCCAGTTGTTCAAAAG
GTTAAAGAGCTAAGGGGAGTTTCTAGTACACCTGGGGTCAACAAGTTTTC
Q F L>
D S P Q R S C G P Q L F K R>
B DOMAIN 3
Q F L D S P Q R S C G P Q L F K R>
CODING REGION
510 520 530 540 550
AGTGAATACTTTGTGTAATGAAAATATCAATGTTGAAAATAATGTAAGCG
TCACTTATGAAACACATTACTTTTATAGTTACAACTTTTATTACATTCGC
V N T L C N E>
B DOMAIN 2
N I N V E N N V S>
A DOMAIN 3
V N T L C N E N I N V E N N V S>
CODING REGION
560 570 580 590 600
TGTCGAAAAGCTGTTGCGAATCAGCGGCAGGATGCACGGATGATTGGATT
ACAGCTTTTCGACAACGCTTAGTCGCCGTCCTACGTGCCTACTAACCTAA
V S K S C C E S A A G C T D D W I>
A DOMAIN 3
V S K S C C E S A A G C T D D W I>
CODING REGION
610 620 630 640 650
AAGAAGAATGTCTGCACACAGCATAAGCCTTTTGTTTTCCGTCCAGGCTT
TTCTTCTTACAGACGTGTGTCGTATTCGGAAAACAAAAGGCAGGTCCGAA
K K N V C T Q H K P F V F R P G F>
A DOMAIN 3
K K N V C T Q H K P F V F R P G F>
CODING REGION
TTACTGA
AATGACT
Y>
Y *>

FIG. 32C

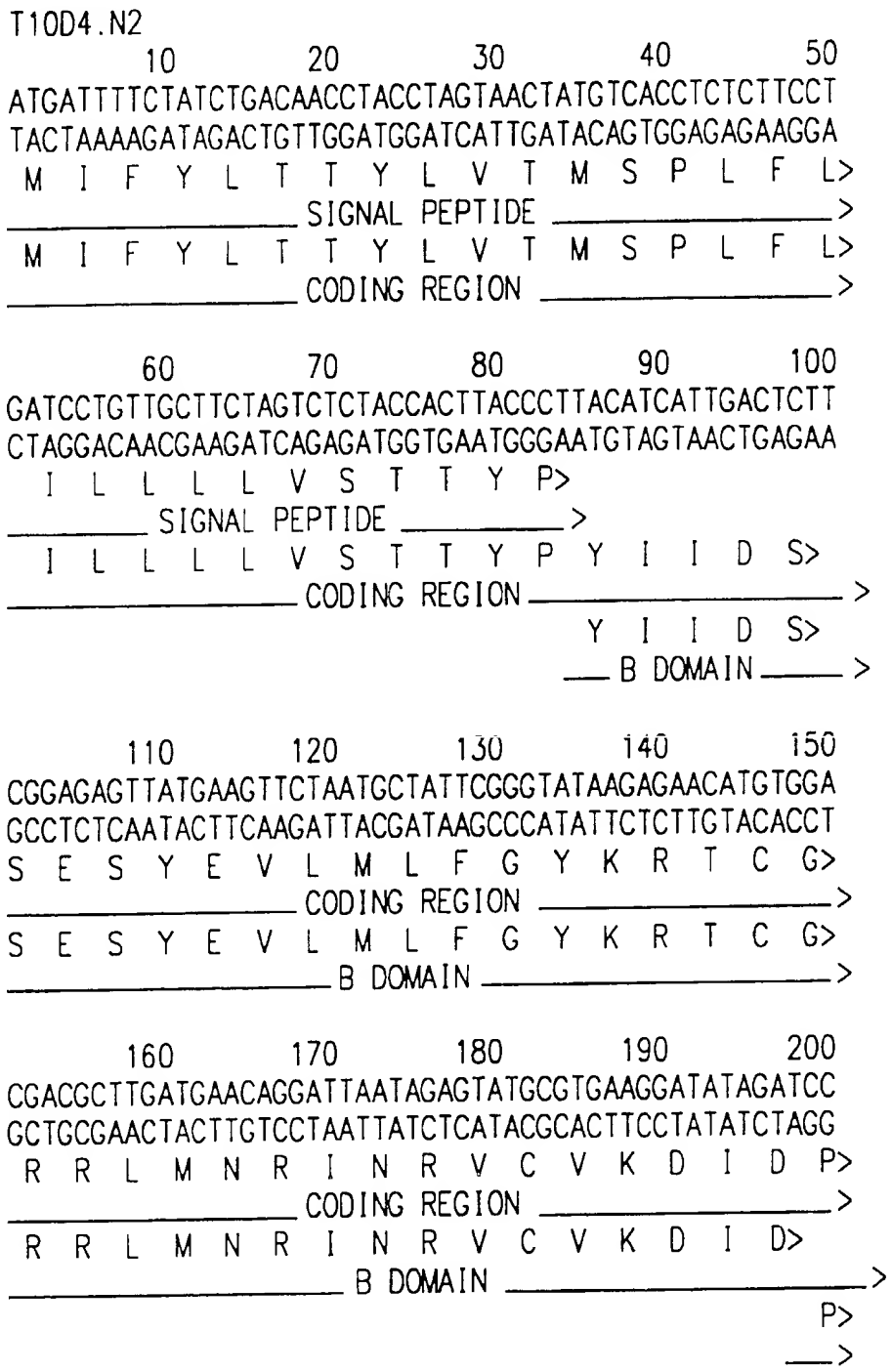

T10D4.N2
10 20 30 40 50
ATGATTTTCTATCTGACAACCTACCTAGTAACTATGTCACCTCTCTTCCT
TACTAAAAGATAGACTGTTGGATGGATCATTGATACAGTGGAGAGAAGGA
M I F Y L T T Y L V T M S P L F L>
SIGNAL PEPTIDE
M I F Y L T T Y L V T M S P L F L>
CODING REGION
60 70 80 90 100
GATCCTGTTGCTTCTAGTCTCTACCACTTACCCTTACATCATTGACTCTT
CTAGGACAACGAAGATCAGAGATGGTGAATGGGAATGTAGTAACTGAGAA
I L L L L V S T T Y P>
SIGNAL PEPTIDE
I L L L L V S T T Y P Y I I D S>
CODING REGION
Y I I D S>
B DOMAIN
110 120 130 140 150
CGGAGAGTTATGAAGTTCTAATGCTATTCGGGTATAAGAGAACATGTGGA
GCCTCTCAATACTTCAAGATTACGATAAGCCCATATTCTCTTGTACACCT
S E S Y E V L M L F G Y K R T C G>
CODING REGION
S E S Y E V L M L F G Y K R T C G>
B DOMAIN
160 170 180 190 200
CGACGCTTGATGAACAGGATTAATAGAGTATGCGTGAAGGATATAGATCC
GCTGCGAACTACTTGTCCTAATTATCTCATACGCACTTCCTATATCTAGG
R R L M N R I N R V C V K D I D P>
CODING REGION
R R L M N R I N R V C V K D I D>
B DOMAIN
P>
>

FIG. 33A

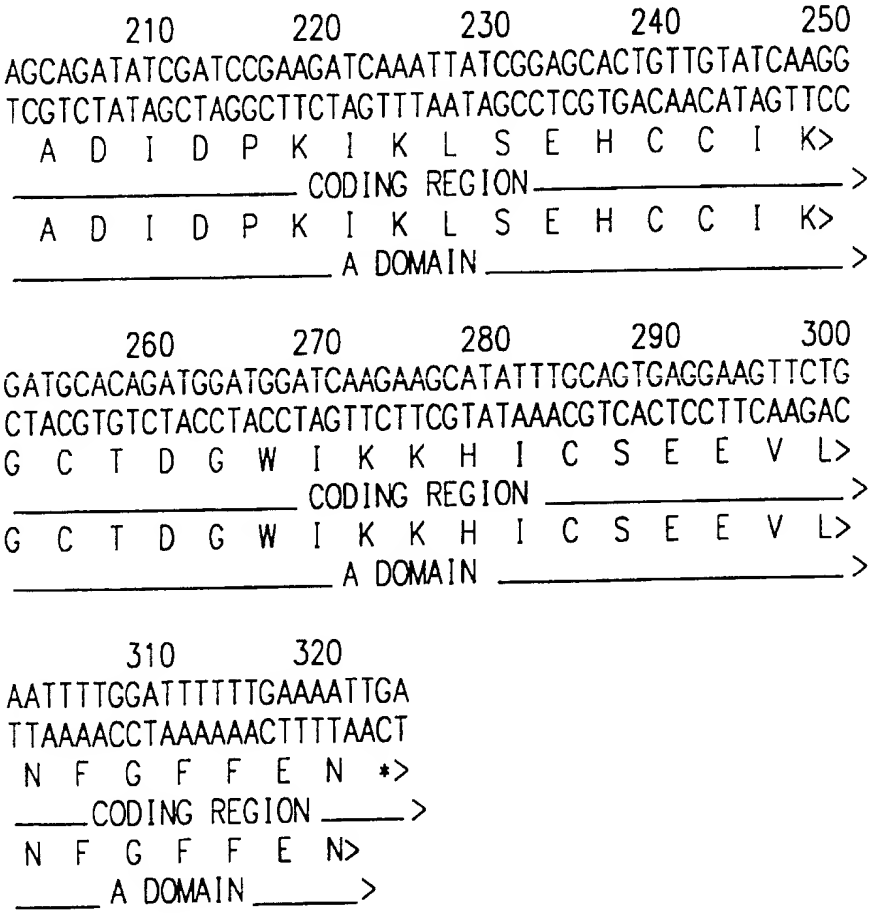

210 220 230 240 250
AGCAGATATCGATCCGAAGATCAAATTATCGGAGCACTGTTGTATCAAGG
TCGTCTATAGCTAGGCTTCTAGTTTAATAGCCTCGTGACAACATAGTTCC
A D I D P K I K L S E H C C I K>
CODING REGION
A D I D P K I K L S E H C C I K>
A DOMAIN
260 270 280 290 300
GATGCACAGATGGATGGATCAAGAAGCATATTTCCAGTGAGGAAGTTCTG
CTACGTGTCTACCTACCTAGTTCTTCGTATAAACGTCACTCCTTCAAGAC
G C T D G W I K K H I C S E E V L>
CODING REGION
G C T D G W I K K H I C S E E V L>
A DOMAIN
310 320
AATTTTGGATTTTTTGAAAATTGA
TTAAAACCTAAAAAACTTTTAACT
N F G F F E N *>
CODING REGION
N F G F F E N>
A DOMAIN

FIG. 33B

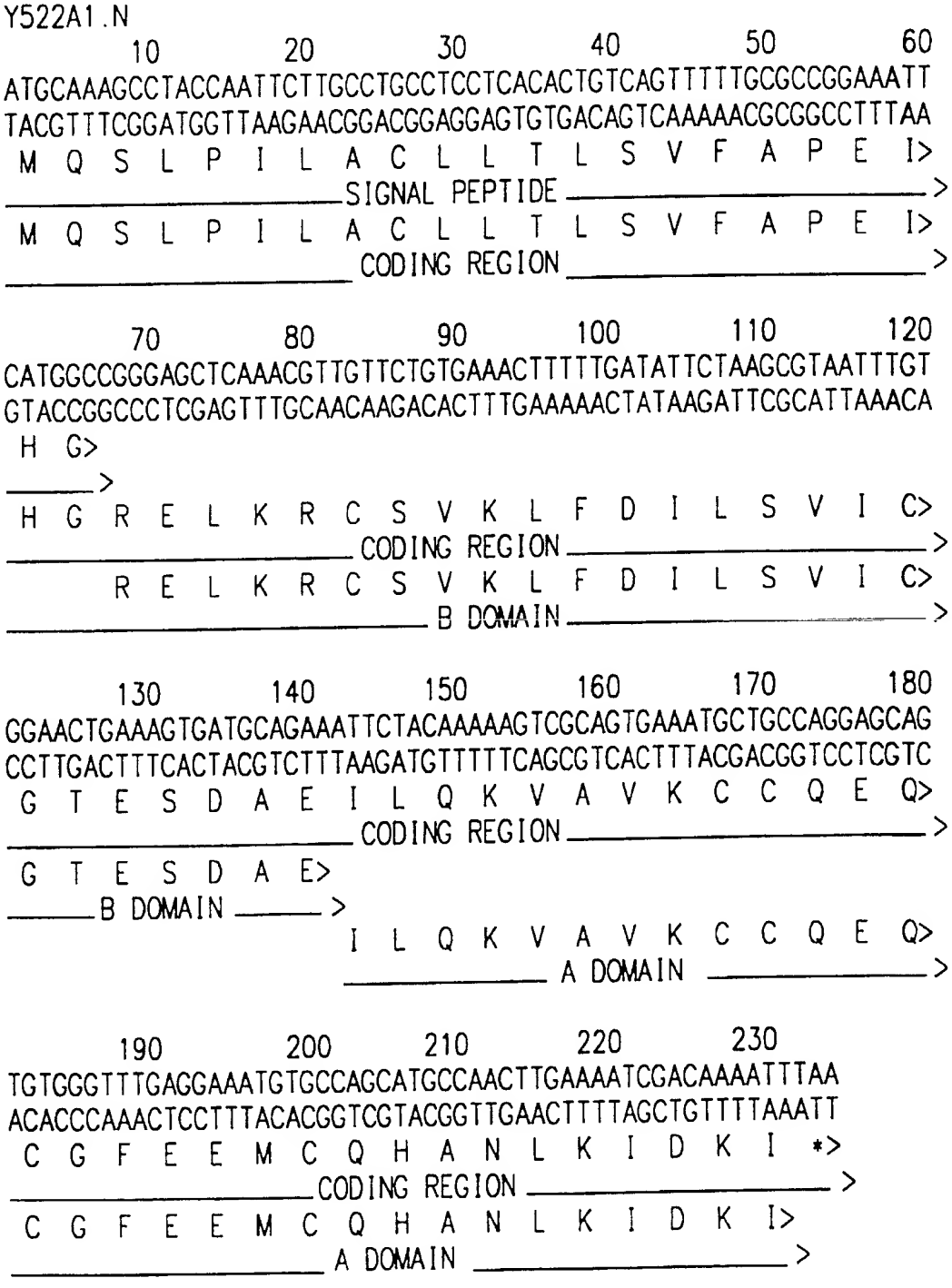

Y522A1.N
10 20 30 40 50 60
ATGCAAAGCCTACCAATTCTTGCCTGCCTCCTCACACTGTCAGTTTTTGCGCCGGAAATT
TACGTTTCGGATGGTTAAGAACGGACGGAGGAGTGTGACAGTCAAAAACGCGGCCTTTAA
M Q S L P I L A C L L T L S V F A P E I>
SIGNAL PEPTIDE
M Q S L P I L A C L L T L S V F A P E I>
CODING REGION
70 80 90 100 110 120
CATGGCCGGGAGCTCAAACGTTGTTCTGTGAAACTTTTTGATATTCTAAGCGTAATTTGT
GTACCGGCCCTCGAGTTTGCAACAAGACACTTTGAAAAACTATAAGATTCGCATTAAACA
H G>
H G R E L K R C S V K L F D I L S V I C>
CODING REGION
R E L K R C S V K L F D I L S V I C>
B DOMAIN
130 140 150 160 170 180
GGAACTGAAAGTGATGCAGAAATTCTACAAAAAGTCGCAGTGAAATGCTGCCAGGAGCAG
CCTTGACTTTCACTACGTCTTTAAGATGTTTTTCAGCGTCACTTTACGACGGTCCTCGTC
G T E S D A E I L Q K V A V K C C Q E Q>
CODING REGION
G T E S D A E>
B DOMAIN
I L Q K V A V K C C Q E Q>
A DOMAIN
190 200 210 220 230
TGTGGGTTTGAGGAAATGTGCCAGCATGCCAACTTGAAAATCGACAAAATTTAA
ACACCCAAACTCCTTTACACGGTCGTACGGTTGAACTTTTAGCTGTTTTAAATT
C G F E E M C Q H A N L K I D K I *>
CODING REGION
C G F E E M C Q H A N L K I D K I>
A DOMAIN

FIG. 34

SEQUENCE LISTING

<110> EXELIXIS PHARMACEUTICALS, INC.

<120> NUCLEIC ACIDS AND PROTEINS OF C. ELEGANS INSULIN-LIKE GENES AND USES THEREOF

<130> 7326-098-228

<140> PCT/US99/
<141> 1999-04-15

<150> 09/062,580
<151> 1998-04-17

<150> 09/074,984
<151> 1998-05-08

<150> 09/084,303
<151> 1998-05-26

<160> 215

<170> PatentIn Ver. 2.0

<210> 1
<211> 109
<212> PRT
<213> Caenorhabditis elegans

<400> 1

```
Met Tyr Trp Phe Arg Gln Val Tyr Arg Pro Ser Phe Phe Phe Gly Phe
 1               5                  10                  15

Leu Ala Ile Leu Leu Leu Ser Ser Pro Thr Pro Ser Asp Ala Ser Ile
            20                  25                  30

Arg Leu Cys Gly Ser Arg Leu Thr Thr Thr Leu Leu Ala Val Cys Arg
        35                  40                  45

Asn Gln Leu Cys Thr Gly Leu Thr Ala Phe Lys Arg Ser Ala Asp Gln
    50                  55                  60

Ser Tyr Ala Pro Thr Thr Arg Asp Leu Phe His Ile His His Gln Gln
65                  70                  75                  80

Lys Arg Gly Gly Ile Ala Thr Glu Cys Cys Glu Lys Arg Cys Ser Phe
                85                  90                  95
```

Ala Tyr Leu Lys Thr Phe Cys Cys Asn Gln Asp Asp Asn
100 105

<210> 2
<211> 91
<212> PRT
<213> Caenorhabditis elegans

<400> 2
Met Ser Ser Tyr Arg Gln Thr Leu Phe Ile Leu Ile Ile Leu Ile Val
1 5 10 15

Ile Ile Leu Phe Val Asn Glu Gly Gln Gly Ala Pro His His Asp Lys
20 25 30

Arg His Thr Ala Cys Val Leu Lys Ile Phe Lys Ala Leu Asn Val Met
35 40 45

Cys Asn His Glu Gly Asp Ala Asp Val Leu Arg Arg Thr Ala Ser Asp
50 55 60

Cys Cys Arg Glu Ser Cys Ser Leu Thr Glu Met Leu Ala Ser Cys Thr
65 70 75 80

Leu Thr Ser Ser Glu Glu Ser Thr Arg Asp Ile
85 90

<210> 3
<211> 106
<212> PRT
<213> Caenorhabditis elegans

<400> 3
Met Phe Ser Phe Phe Thr Tyr Phe Leu Leu Ser Ala Leu Leu Leu Ser
1 5 10 15

Ala Ser Cys Arg Gln Pro Ser Met Asp Thr Ser Lys Ala Asp Arg Ile
20 25 30

Leu Arg Glu Ile Glu Met Glu Thr Glu Leu Glu Asn Gln Leu Ser Arg
35 40 45

Ala Arg Arg Val Pro Ala Gly Glu Val Arg Ala Cys Gly Arg Arg Leu
50 55 60

```
Leu Leu Phe Val Trp Ser Thr Cys Gly Glu Pro Cys Thr Pro Gln Glu
 65                  70                  75                  80

Asp Met Asp Ile Ala Thr Val Cys Cys Thr Thr Gln Cys Thr Pro Ser
                 85                  90                  95

Tyr Ile Lys Gln Ala Cys Cys Pro Glu Lys
            100                 105
```

<210> 4
<211> 106
<212> PRT
<213> Caenorhabditis elegans

<400> 4

```
Met Asn Ala Ile Ile Phe Cys Leu Leu Phe Thr Thr Val Thr Ala Thr
  1               5                  10                  15

Tyr Glu Val Phe Gly Lys Gly Ile Glu His Arg Asn Glu His Leu Ile
            20                  25                  30

Ile Asn Gln Leu Asp Ile Ile Pro Val Glu Ser Thr Pro Thr Pro Asn
         35                  40                  45

Arg Ala Ser Arg Val Gln Lys Arg Leu Cys Gly Arg Arg Leu Ile Leu
     50                  55                  60

Phe Met Leu Ala Thr Cys Gly Glu Cys Asp Thr Asp Ser Ser Glu Asp
 65                  70                  75                  80

Leu Ser His Ile Cys Cys Ile Lys Gln Cys Asp Val Gln Asp Ile Ile
                 85                  90                  95

Arg Val Cys Cys Pro Asn Ser Phe Arg Lys
            100                 105
```

<210> 5
<211> 107
<212> PRT
<213> Caenorhabditis elegans

<400> 5

```
Met Lys Leu Ser Val Val Leu Ala Leu Phe Ile Ile Phe Gln Leu Gly
  1               5                  10                  15

Ala Ala Ser Leu Met Arg Asn Trp Met Phe Asp Phe Glu Lys Glu Leu
```

```
            20                  25                  30

Glu His Asp Tyr Asp Asp Ser Glu Ile Gly Phe His Asn Ile His Ser
        35                  40                  45

Leu Met Ala Arg Ser Arg Arg Gly Asp Lys Val Lys Ile Cys Gly Thr
    50                  55                  60

Lys Val Leu Lys Met Val Met Val Met Cys Gly Gly Glu Cys Ser Ser
65                  70                  75                  80

Thr Asn Glu Asn Ile Ala Thr Glu Cys Cys Glu Lys Met Cys Thr Met
                85                  90                  95

Glu Asp Ile Thr Thr Lys Cys Cys Pro Ser Arg
            100                 105
```

<210> 6
<211> 112
<212> PRT
<213> Caenorhabditis elegans

<400> 6

```
Met Asn Ser Val Phe Thr Ile Ile Phe Val Leu Cys Ala Leu Gln Val
  1               5                  10                  15

Ala Ala Ser Phe Arg Gln Ser Phe Gly Pro Ser Met Ser Glu Glu Ser
            20                  25                  30

Ala Ser Met Gln Leu Leu Arg Glu Leu Gln His Asn Met Met Glu Ser
        35                  40                  45

Ala His Arg Pro Met Pro Arg Ala Arg Arg Val Pro Ala Pro Gly Glu
    50                  55                  60

Thr Arg Ala Cys Gly Arg Lys Leu Ile Ser Leu Val Met Ala Val Cys
65                  70                  75                  80

Gly Asp Leu Cys Asn Pro Gln Glu Gly Lys Asp Ile Ala Thr Glu Cys
                85                  90                  95

Cys Gly Asn Gln Cys Ser Asp Asp Tyr Ile Arg Ser Ala Cys Cys Pro
            100                 105                 110
```

<210> 7
<211> 100
<212> PRT
<213> Caenorhabditis elegans

<400> 7
Met His Ser Ile Val Ala Leu Met Leu Ile Gly Thr Ile Leu Pro Ile
1 5 10 15

Ala Ala Leu His Gln Lys His Gln Gly Phe Ile Leu Ser Ser Ser Asp
20 25 30

Ser Thr Gly Asn Gln Pro Met Asp Ala Ile Ser Arg Ala Asp Arg His
35 40 45

Thr Asn Tyr Arg Ser Cys Ala Leu Arg Leu Ile Pro His Val Trp Ser
50 55 60

Val Cys Gly Asp Ala Cys Gln Pro Gln Asn Gly Ile Asp Val Ala Gln
65 70 75 80

Lys Cys Cys Ser Thr Asp Cys Ser Ser Asp Tyr Ile Lys Glu Ile Cys
85 90 95

Cys Pro Phe Asp
100

<210> 8
<211> 105
<212> PRT
<213> Caenorhabditis elegans

<400> 8
Met Pro Pro Ile Ile Leu Val Phe Phe Leu Val Leu Ile Pro Ala Ser
1 5 10 15

Gln Gln Tyr Pro Phe Ser Leu Glu Ser Leu Asn Asp Gln Ile Ile Asn
20 25 30

Glu Glu Val Ile Glu Tyr Met Leu Glu Asn Ser Ile Arg Ser Ser Arg
35 40 45

Thr Arg Arg Val Pro Asp Glu Lys Lys Ile Tyr Arg Cys Gly Arg Arg
50 55 60

Ile His Ser Tyr Val Phe Ala Val Cys Gly Lys Ala Cys Glu Ser Asn

65 70 75 80

Thr Glu Val Asn Ile Ala Ser Lys Cys Cys Arg Glu Glu Cys Thr Asp
85 90 95

Asp Phe Ile Arg Lys Gln Cys Cys Pro
100 105

<210> 9
<211> 104
<212> PRT
<213> Caenorhabditis elegans

<400> 9
Met Ser Pro Ile Ile Leu Ile Phe Phe Leu Val Phe Ile Pro Phe Ser
1 5 10 15

Gln Gln His Thr Ser Leu Glu Glu Ser Leu Asn Asp Arg Ile Ile Ser
20 25 30

Glu Glu Val Val Glu Met Leu Ser Glu Lys Glu Ile Arg Pro Ser Arg
35 40 45

Val Arg Arg Val Pro Glu Gln Lys Asn Lys Leu Cys Gly Lys Gln Val
50 55 60

Leu Ser Tyr Val Met Ala Leu Cys Glu Lys Ala Cys Asp Ser Asn Thr
65 70 75 80

Lys Val Asp Ile Ala Thr Lys Cys Cys Arg Asp Ala Cys Ser Asp Glu
85 90 95

Phe Ile Arg His Gln Cys Cys Pro
100

<210> 10
<211> 118
<212> PRT
<213> Caenorhabditis elegans

<400> 10
Met Ile Val Thr Leu Ile Val Phe Leu Val Ile Gly Leu Gln Met Ala
1 5 10 15

His Leu Ser Gln Val Ser Gly Asn Asn Glu Asn Gly Phe Leu Asn Pro
20 25 30

```
Phe Asp Leu Ser Gln Trp Ser Glu Glu Ile Leu His Arg Gln Tyr His
          35                  40                  45

His His His His His His His Gly Asn Arg Ala Arg Arg Thr Leu Glu
      50                  55                  60

Thr Glu Lys Ile Tyr Arg Cys Gly Arg Lys Leu Tyr Thr Asp Val Leu
 65                  70                  75                  80

Ser Ala Cys Asn Gly Pro Cys Glu Pro Gly Thr Glu Gln Asp Leu Ser
                 85                  90                  95

Lys Leu Cys Cys Gly Asn Gln Cys Thr Phe Val Glu Ile Arg Lys Ala
            100                 105                 110

Cys Cys Ala Asp Lys Leu
        115
```

<210> 11
<211> 86
<212> PRT
<213> Caenorhabditis elegans

<400> 11

```
Met Gln Ser Asn Ile Thr Ala Ser Leu Phe Ile Ala Leu Leu Ile Phe
  1               5                  10                  15

Gly Val Ile Ser Ala Ala Pro Ser His Glu Lys Thr His Lys Lys Cys
            20                  25                  30

Ser Asp Lys Leu Tyr Leu Ala Met Lys Ser Leu Cys Ser Tyr Arg Gly
          35                  40                  45

Tyr Ser Glu Phe Leu Arg Asn Ser Ala Thr Lys Cys Cys Gln Asp Asn
      50                  55                  60

Cys Glu Ile Ser Glu Met Met Ala Leu Cys Val Val Ala Pro Asn Phe
 65                  70                  75                  80

Asp Asp Asp Leu Leu His
                 85
```

<210> 12
<211> 76
<212> PRT

<213> Caenorhabditis elegans

<400> 12

```
Met Lys Thr Tyr Ser Phe Phe Val Leu Phe Ile Val Phe Ile Phe Phe
  1               5                  10                  15

Ile Ser Ser Ser Lys Ser His Ser Lys Lys His Val Arg Phe Leu Cys
             20                  25                  30

Ala Thr Lys Ala Val Lys His Ile Arg Lys Val Cys Pro Asp Met Cys
         35                  40                  45

Leu Thr Gly Glu Glu Val Glu Val Asn Glu Phe Cys Lys Met Gly Tyr
     50                  55                  60

Ser Asp Ser Gln Ile Lys Tyr Ile Cys Cys Pro Glu
 65                  70                  75
```

<210> 13
<211> 83
<212> PRT
<213> Caenorhabditis elegans

<400> 13

```
Met His Thr Thr Thr Ile Leu Ile Cys Phe Phe Ile Phe Leu Val Gln
  1               5                  10                  15

Val Ser Thr Met Asp Ala His Thr Asp Lys Tyr Val Arg Thr Leu Cys
             20                  25                  30

Gly Lys Thr Ala Ile Arg Asn Ile Ala Asn Leu Cys Pro Pro Lys Pro
         35                  40                  45

Glu Met Lys Gly Ile Cys Ser Thr Gly Glu Tyr Pro Ser Ile Thr Glu
     50                  55                  60

Tyr Cys Ser Met Gly Phe Ser Asp Ser Gln Ile Lys Phe Met Cys Cys
 65                  70                  75                  80

Asp Asn Gln
```

<210> 14
<211> 76
<212> PRT
<213> Caenorhabditis elegans

<400> 14

```
Met Phe Val Leu Leu Ile Ile Leu Ser Ile Ile Leu Ala Gln Val Thr
  1               5                  10                  15

Asp Ala His Ser Glu Leu His Val Arg Arg Val Cys Gly Thr Ala Ile
             20                  25                  30

Ile Lys Asn Ile Met Arg Leu Cys Pro Gly Val Pro Ala Cys Glu Asn
         35                  40                  45

Gly Glu Val Pro Ser Pro Thr Glu Tyr Cys Ser Met Gly Tyr Ser Asp
     50                  55                  60

Ser Gln Val Lys Tyr Leu Cys Cys Pro Thr Ser Gln
 65                  70                  75
```

<210> 15
<211> 80
<212> PRT
<213> Caenorhabditis elegans

<400> 15

```
Met Asp Lys Pro Ser Tyr Leu Ser Ser Lys Glu Ala Trp Lys Met Leu
  1               5                  10                  15

Asn Glu Leu Leu Lys Glu Pro Lys His His His His His His Arg His
             20                  25                  30

Lys Gly Tyr Cys Gly Val Lys Ala Val Lys Lys Leu Lys Gln Ile Cys
         35                  40                  45

Pro Asp Leu Cys Ser Asn Val Asp Asp Asn Leu Leu Met Glu Met Cys
     50                  55                  60

Ser Lys Asn Leu Thr Asp Asp Asp Ile Leu Gln Arg Cys Cys Pro Glu
 65                  70                  75                  80
```

<210> 16
<211> 108
<212> PRT
<213> Caenorhabditis elegans

<400> 16

```
Met Phe Ser Thr Arg Gly Val Leu Leu Leu Ser Leu Met Ala Ala
 1               5                  10                  15

Val Ala Ala Phe Gly Leu Phe Ser Arg Pro Ala Pro Ile Thr Arg Asp
            20                  25                  30

Thr Ile Arg Pro Pro Arg Ala Lys His Gly Ser Leu Lys Leu Cys Pro
        35                  40                  45

Pro Gly Gly Ala Ser Phe Leu Asp Ala Phe Asn Leu Ile Cys Pro Met
    50                  55                  60

Arg Arg Arg Arg Arg Ser Val Ser Glu Asn Tyr Asn Asp Gly Gly Gly
65                  70                  75                  80

Ser Leu Leu Gly Arg Thr Met Asn Met Cys Cys Glu Thr Gly Cys Glu
                85                  90                  95

Phe Thr Asp Ile Phe Ala Ile Cys Asn Pro Phe Gly
            100                 105
```

<210> 17
<211> 95
<212> PRT
<213> Caenorhabditis elegans

<400> 17

```
Met Val His Arg Leu Phe Ile Val Leu Ile Ala Ile Ile Leu Val Ala
 1               5                  10                  15

Lys Ser Thr Ala Ile Ser Leu Gln Gln Ala Asp Gly Arg Met Lys Met
            20                  25                  30

Cys Pro Pro Gly Gly Ser Thr Phe Thr Met Ala Trp Ser Met Ser Cys
        35                  40                  45

Ser Met Arg Arg Arg Lys Arg Asp Val Gly Arg Tyr Phe Glu Lys Arg
    50                  55                  60

Ala Leu Ile Ala Pro Ser Ile Arg Gln Leu Gln Thr Ile Cys Cys Gln
65                  70                  75                  80

Val Gly Cys Asn Val Glu Asp Leu Leu Ala Tyr Cys Ala Pro Ile
                85                  90                  95
```

<210> 18
<211> 83
<212> PRT
<213> Caenorhabditis elegans

<400> 18

```
Met Lys Phe Phe Arg Leu Ile Leu Leu Cys Ala Leu Val Leu Thr Thr
  1               5                  10                  15

Met Ala Phe Leu Ala Pro Ser Thr Ala Ala Lys Arg Arg Cys Gly Arg
             20                  25                  30

Arg Leu Ile Pro Tyr Val Tyr Ser Ile Cys Gly Gly Pro Cys Glu Asn
         35                  40                  45

Gly Asp Ile Ile Ile Glu His Cys Phe Ser Gly Thr Thr Pro Thr Ile
     50                  55                  60

Ala Glu Val Gln Lys Ala Cys Cys Pro Glu Leu Ser Glu Asp Pro Thr
 65                  70                  75                  80

Phe Ser Ser
```

<210> 19
<211> 321
<212> DNA
<213> Caenorhabditis elegans

<400> 19

```
atgttttcat tctttacata tttccttctc tccgcacttc ttctctccgc ttcatgtcga 60
caaccttcca tggacaccag caaagccgat cgtattctac gagagatcga aatggaaaca 120
gaactcgaaa atcaactctc ccgagcacga cgagtcccag ctggagaggt tcgtgcctgt 180
ggaagacgac ttcttctctt tgtctggtca acctgtggag aaccatgcac gccacaagag 240
gacatggaca ttgccacagt ttgctgcaca acacagtgca ctccatcata tataaaacaa 300
gcttgctgcc cagaaaagta a                                          321
```

<210> 20
<211> 321
<212> DNA
<213> Caenorhabditis elegans

<400> 20

```
atgttttcat tctttacata tttccttctc tccgcacttc ttctctccgc ttcatgtcga 60
caaccttcca tggacaccag caaagccgat cgtattctac gagagatcga aatggaaaca 120
gaactcgaaa atcaactctc ccgagcacga cgagtcccag ctggagaggt tcgtgcctgt 180
ggaagacgac ttcttctctt tgtctggtca acctgtggag aaccatgcac gccacaagag 240
```

gacatggaca ttgccacagt ttgctgcaca acacagtgca ctccatcata tataaaacaa 300
gcttgctgcc cagaaaagta a 321

<210> 21
<211> 321
<212> DNA
<213> Caenorhabditis elegans

<400> 21
atgaacgcta taatcttctg tctcctcttc acaactgtca ctgccactta tgaagttttc 60
ggaaaaggaa tagaacacag aaatgaacat ttgatcatca atcaacttga tatcatacca 120
gttgagtcaa ctccaactcc aaaccgtgcc tcaagagtcc agaaacgtct atgcggaaga 180
cgtcttattt tattcatgct tgcaacatgt ggagaatgtg atacagattc atcagaagac 240
ctttcgcata tttgctgcat aaaacaatgt gacgttcaag atatcatcag agtctgctgc 300
ccgaattcat ttagaaaata g 321

<210> 22
<211> 324
<212> DNA
<213> Caenorhabditis elegans

<400> 22
atgaaactct ccgttgttct tgcacttttc attattttcc aacttggagc tgcaagtctt 60
atgcgtaact ggatgttcga ttttgagaaa gaattggaac acgattatga tgattcggaa 120
attggattcc ataacattca ctccctgatg gccagatcaa gaagaggaga caaagtgaag 180
atttgtggta caaaagttct gaaaatggtg atggtaatgt gtggaggaga atgttcatca 240
acgaatgaga acatcgctac agaatgctgt gaaaaaatgt gcacaatgga agatataact 300
actaagtgct gcccttcaag atga 324

<210> 23
<211> 339
<212> DNA
<213> Caenorhabditis elegans

<400> 23
atgaactctg tctttactat catcttcgtt ttgtgcgcac tccaagtcgc tgcaagtttc 60
cgtcaatcct tcggtccttc aatgtctgaa gaatcagcaa gcatgcaact tctccgtgaa 120
cttcaacaca acatgatgga atcagctcac cgaccaatgc cacgagcaag acgtgttcca 180
gcaccaggag aaactcgtgc ctgcggaaga aaactcatct ctttagtcat ggctgtctgt 240
ggagatcttt gcaacccaca agaaggaaag gacattgcga ctgaatgctg cggaaatcag 300
tgttctgatg actacataag atctgcttgt tgtccatga 339

<210> 24
<211> 303
<212> DNA
<213> Caenorhabditis elegans

<400> 24

atgcactcga tcgtcgcctt gatgctcatc ggaacaattc tcccaatcgc tgctcttcac 60
cagaagcatc aaggcttcat cctgtcgtca tccgattcaa ccggaaacca accaatggat 120
gcgatctcaa gagccgaccg tcacaccaac taccgatcat gcgcattgcg gctcatcccg 180
catgtctggt cggtgtgcgg tgacgcctgc caaccacaaa acggaatcga tgtcgctcaa 240
aaatgttgct ccactgattg cagctccgat tacatcaaag aaatctgctg cccatttgac 300
taa 303

<210> 25
<211> 318
<212> DNA
<213> Caenorhabditis elegans

<400> 25
atgccaccaa taattttggt tttcttttg gttttaatcc ctgcttctca acaatatcct 60
ttttcactgg agtccttaaa tgatcaaata atcaatgaag aagtaatcga atatatgctt 120
gaaaattcaa ttaggtccag cagaaccaga agagtccctg acgagaaaaa aatttatcgt 180
tgtggaagaa gaatacattc gtatgtgttt gcggtttgtg gaaaagcatg cgaatcgaat 240
actgaagtta atattgcatc aaaatgttgc cgtgaagaat gcaccgacga cttcattcga 300
aaacagtgct gtccttaa 318

<210> 26
<211> 315
<212> DNA
<213> Caenorhabditis elegans

<400> 26
atgtcgccaa tcattttgat tttcttttg gttttcattc cgttttctca acaacacaca 60
tctttagagg agtccttaaa tgatcgaata atcagtgaag aagtagtcga aatgctatca 120
gagaaagaaa ttagacccag cagagtaaga agagtccctg aacaaaaaaa taaattgtgc 180
ggaaagcaag tcttatccta cgttatggca ctttgtgaaa aagcatgcga ttcaaataca 240
aaagtcgata ttgcgacaaa atgttgccgc gatgcatgct cagacgaatt cattcgacat 300
caatgttgtc cttaa 315

<210> 27
<211> 357
<212> DNA
<213> Caenorhabditis elegans

<400> 27
atgatcgtca ctttgattgt ctttcttgtc attggacttc aaatggcaca cctttctcaa 60
gtatctggaa acaacgaaaa tggattctta aatccatttg atttgtctca atggagcgaa 120
gaaatcctcc accgtcagta tcatcatcac caccaccatc accatggaaa tcgggcgaga 180
agaaccttgg aaaccgaaaa aatctaccgc tgtggaagaa aactctacac tgatgtgcta 240
tcagcgtgca acgggccatg tgaaccgggt acggaacagg atctctctaa gctgtgctgt 300
ggaaaccaat gtactttcgt tgaaatcagg aaagcatgct gtgccgacaa attgtaa 357

<210> 28
<211> 276

<212> DNA
<213> Caenorhabditis elegans

<400> 28
atgtctagtt accgtcaaac attgttcatt cttattattc ttattgtaat tattctcttc 60
gtcaatgagg gtcaaggagc gcctcaccat gacaaacggc acactgcatg cgtcctaaag 120
attttcaagg cgctaaacgt tatgtgtaat catgaaggtg atgcagatgt tctgaggaga 180
acagcatccg actgctgtcg ggagagctgc tcgctaacag aaatgttagc gagctgcacc 240
ctcaccagct cagaagagtc aactcgggac atttaa 276

<210> 29
<211> 261
<212> DNA
<213> Caenorhabditis elegans

<400> 29
atgcaatcaa acatcaccgc ttcattattc atagcgttgc ttatatttgg agtaatcagt 60
gcagctccat ctcatgaaaa aacacacaaa aaatgctctg ataaattata tttggcgatg 120
aagtcgttgt gtagttatcg aggttatagt gaattcttaa gaaattctgc aactaagtgt 180
tgccaagaca attgtgagat ttcggaaatg atggcgttgt gtgttgttgc tcccaatttt 240
gacgacgatc tccttcatta a 261

<210> 30
<211> 231
<212> DNA
<213> Caenorhabditis elegans

<400> 30
atgaaaacct actcattttt cgtgcttttt attgtattca tctttttttat ttcttcatca 60
aaatctcatt caaagaaaca tgttcgtttc ctttgtgcaa caaaagcggt caaacacatt 120
cggaaagtat gccctgatat gtgtctcact ggagaagaag tcgaagtcaa tgagttttgc 180
aagatggggt actcggattc tcaaatcaag tacatttgct gtcccgaata a 231

<210> 31
<211> 252
<212> DNA
<213> Caenorhabditis elegans

<400> 31
atgcacacta caactattct catatgcttt ttcatctttc ttgttcaagt ctccacaatg 60
gatgctcaca ctgacaaata cgtcagaact ctgtgtggaa aaactgcaat cagaaatatt 120
gccaaccttt gcccgccaaa gccagaaatg aagggtatct gttctaccgg agagtatcca 180
agcatcaccg aatactgttc catgggattt tcagactctc agatcaagtt tatgtgctgt 240
gataaccaat ga 252

<210> 32
<211> 231
<212> DNA

<213> Caenorhabditis elegans

<400> 32
atgttcgttc ttcttattat tctctctatc attctggctc aagtcactga tgctcattca 60
gagcttcacg ttcgtagggt gtgcggaact gctatcataa agaacataat gcgattgtgc 120
ccaggggtac cggcttgcga aaatggagaa gttccaagtc caaccgagta ctgttcaatg 180
gggtactcag acagccaggt aaaataccta tgctgtccaa cttctcagtg a 231

<210> 33
<211> 243
<212> DNA
<213> Caenorhabditis elegans

<400> 33
atggacaaac catcctacct gtcatccaaa gaagcatgga aaatgctaaa tgagctgctg 60
aaagagccga aacatcatca tcatcatcac aggcacaaag gatattgtgg agttaaagct 120
gtaaagaaat taaaacaaat ctgtccagat ctttgctcga atgttgatga taaccttctc 180
atggaaatgt gctcaaaaaa cctgacggat gatgatattt tgcaacggtg ctgtccagaa 240
tga 243

<210> 34
<211> 327
<212> DNA
<213> Caenorhabditis elegans

<400> 34
atgttctcga ccagaggggt actcctttta ctgtctttga tggctgctgt agccgcattc 60
gggctgtttt ctagaccggc tccaatcact cgggacacta tccgaccacc acgtgccaaa 120
cacggttcgc tgaaattatg cccaccaggt ggtgcctcat tccttgacgc tttcaacttg 180
atttgcccaa tgcgccgtcg acgcaggagt gtttcagaaa actacaacga cggcggtggc 240
agccttttgg gacggacaat gaatatgtgc tgtgagacgg gatgtgaatt cactgacatt 300
ttcgcaatct gcaatccttt tggataa 327

<210> 35
<211> 288
<212> DNA
<213> Caenorhabditis elegans

<400> 35
atggtccacc gacttttcat cgtccttatt gcaattattc ttgtcgcaaa atcaactgca 60
atctcacttc aacaagctga cggacgcatg aaaatgtgcc caccaggtgg ttcaacattc 120
acaatggcat ggtcaatgtc gtgttcgatg cgcaggagaa aacgagatgt tggacgatat 180
ttcgaaaaac gtgctctgat cgccccatca atccgtcaac ttcaaacaat ttgctgtcaa 240
gttggttgca acgtggaaga tcttcttgcc tactgtgccc caatttaa 288

<210> 36
<211> 252
<212> DNA

<213> Caenorhabditis elegans

<400> 36
atgaaattct tccgcttaat cttgctctgc gcccttgtcc tgaccaccat ggctttttg 60
gctccaagta cggcagccaa gaggcgttgt ggccgccgct taattcccta tgtctattca 120
atatgcggcg gcccgtgcga gaatggagat attatcatcg agcactgctt ctccggaaca 180
actcccacca ttgccgaagt ccaaaaggct tgctgtcctg aactatctga agacccaact 240
ttctcatctt aa 252

<210> 37
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 37
gacggagatg gcttgttgga cgac 24

<210> 38
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 38
ggtttaatta cccaagtttg ag 22

<210> 39
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 39
caagagaatg ttttcattct ttac 24

<210> 40
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 40
ttacttttct gggcagcaag cttg 24

<210> 41
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 41
ctaccatgaa cgctataatc ttct 24

<210> 42
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 42
atgatagtac gatatgtcca taac 24

<210> 43
<211> 25
<212> DNA
<213> Caenorhabditis elegans

<400> 43
cctatttcc agccacagca ctctc 25

<210> 44
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 44
ccccgtactc attttccgtt atcc 24

<210> 45
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 45
gtatggtaca gagactgata tcgg 24

<210> 46
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 46
caaggaaaat gcactcgatc gtcg 24

<210> 47
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 47
cccaagcttt gttatttaat gatgtggaga tgg 33

<210> 48
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 48
gctctagaat ggtaaataca gaacattggt tc 32

<210> 49
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 49
gctctagagt gacggtaggt gtgtagatga ac 32

<210> 50
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 50
atcgaaactc ttcaatcttc aagg 24

<210> 51
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 51
gatagaagaa attaaggaca gcac 24

<210> 52
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 52
gtaaacgatt agattaagga caac 24

<210> 53
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 53
gaggagtgaa acgatgatcg tcac 24

<210> 54
<211> 24
<212> DNA
<213> Caenorhabditis elegans

<400> 54
atccaattga gaagacgatt gttg 24

<210> 55
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 55
cccaagcttt tgaaccatga aaacctactc att 33

<210> 56
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 56
gctctagagc ttttttttat tcgggacagc aa 32

<210> 57
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 57
cccaagcttg gatttctgga atttcgataa tg 32

<210> 58
<211> 31
<212> DNA
<213> Caenorhabditis elegans

<400> 58
gctctagagc agcatagaat ggcggaagat c 31

<210> 59
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 59
cccaagcttg tgtaggaatc gttaaatatg tct 33

<210> 60
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 60
gctctagaga gatcatatta tattacacga ac 32

<210> 61
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 61
cccaagcttc cgctctcaac aacgggccac acg 33

<210> 62
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 62
gctctagaga tgaataagtt atcaattatc gt 32

<210> 63
<211> 31
<212> DNA
<213> Caenorhabditis elegans

<400> 63
cccaagcttg gtttaattac ccaagtttga g 31

<210> 64
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 64
gctctagatg atgcgtattt tgtgggcggt ac 32

<210> 65
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 65
gctctagact catcagttga aaatgaattt aag 33

<210> 66
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 66
cccaagcttg gcataagcga gtatctgtga tcc 33

<210> 67
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 67
ccgctcgagg taaagcgagg gtaaagtaga tcg 33

<210> 68
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 68
cccaagcttc taaccaacaa aaatgcacac tac 33

<210> 69
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 69
gctctagaca cgtgaacaat ctttatcttt at 32

<210> 70
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 70
cccaagcttc acagccaaaa acaaaaatgc aatc 34

<210> 71
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 71
gctctagaca cagtatttta atgaaggaga tc 32

<210> 72
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 72
ttgggcgcgc cgtcttgcat gcagttgtca cg 32

<210> 73
<211> 35
<212> DNA
<213> Caenorhabditis elegans

<400> 73
ccaaccggta tcattgcgta ctgtcgtagc gtgtg 35

<210> 74
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 74
ttgggcgcgc ctgctaccgt gggaatttta caag 34

<210> 75
<211> 35
<212> DNA
<213> Caenorhabditis elegans

<400> 75
ccaaccggta tcatggtaga ttttagaatg gaaag 35

<210> 76
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 76
ttgggcgcgc cggagttcat ctggaggtca catc 34

<210> 77
<211> 39
<212> DNA
<213> Caenorhabditis elegans

<400> 77
ccaaccggta tcattattca gaacaggaat tgataaatg 39

<210> 78
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 78
ttgggcgcca gataaataca gaatgggcgg ag 32

<210> 79
<211> 36
<212> DNA
<213> Caenorhabditis elegans

<400> 79
ccaaccggta tcattctctt ggagcttttg aaaaac 36

<210> 80
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 80
ttgggcgcgc cagtcgtcca acaagccatc tcc 33

<210> 81
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 81
ccaaccggtt gcattttcct tgaagattga ag 32

<210> 82
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 82
ttgggcgcgc ctagattttc tccattcaca aac 33

<210> 83
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 83
ccaaccggta tcattataat gatatggata acgg 34

<210> 84
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 84
ttgggcgcgc caatcgtttt catcattttg cttc 34

<210> 85
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 85
ccaaccggta tcatctggaa aagtaatatt atat 34

<210> 86
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 86
ttgggcgcgc ctgaaatctt tatatcctct tcac 34

<210> 87
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 87
ccaaccggta tcatctggaa ataattaata tcag 34

<210> 88
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 88
ttgggcgcgc ctaacacgtg cattggaggc ggag 34

<210> 89
<211> 36
<212> DNA
<213> Caenorhabditis elegans

<400> 89
ccaacggtat catcgtttca ctcctcgaat tatttg 36

<210> 90
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 90
ttgggcgcgc cattggtatc acaaggatca agc 33

<210> 91
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 91
ccaaccggca tttttgtttt tggctgtgat ta 32

<210> 92
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 92
ttgggcgcgc caattttgac gacgatcttc ttc 33

<210> 93
<211> 37
<212> DNA
<213> Caenorhabditis elegans

<400> 93
ccaaccggta tcatatttaa cgattcctac acaaacc 37

<210> 94
<211> 30
<212> DNA
<213> Caenorhabditis elegans

<400> 94
ttgggcgcgc cgtgtggagg tggtgaatcc 30

<210> 95
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 95
cggggtaccc tcatttcaaa gaaatgttga ata 33

<210> 96
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 96
ttgggcgcgc cggagccgaa caagaaaaac ctac 34

<210> 97
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 97
ccaaccggtt tcatggttca actcaaaaag gaa 33

<210> 98
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 98
ttgggcgcgc cagttcgtct cagcatcatc ttgc 34

<210> 99
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 99
ccaaccggtt tcatggttca actcaaaaag gaa 33

<210> 100
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 100
ttgggcgcgc catgggattt tcagactctc ag 32

<210> 101
<211> 35
<212> DNA
<213> Caenorhabditis elegans

<400> 101
ccaaccggta acattatcga aattccagaa atccg 35

<210> 102
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 102
ttgggcgcgc cacttcggac agatgtgaca cg 32

<210> 103
<211> 35
<212> DNA
<213> Caenorhabditis elegans

<400> 103
cggggtacct gcattgtaaa agtgattttg aaaat 35

<210> 104
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 104
caaacagttg tagctcaaag gc 22

<210> 105
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 105
gcatacggta cctatrcgtt tc 22

<210> 106
<211> 21
<212> DNA
<213> Caenorhabditis elegans

<400> 106
agctcaaagg ccaaatgtgt g 21

<210> 107
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 107
aacaaaccct acagttactg gg 22

<210> 108
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 108
gctatccacc tgtccaacct ac 22

<210> 109
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 109
ggaggctctt tactcgcctt ac 22

<210> 110
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 110
tacaggctgt ccttctgtta cg 22

<210> 111
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 111
tccactattc cggtaatacc tc 22

<210> 112
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 112
gtaagaaatc gagagtcacg cc 22

<210> 113
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 113
ctgcctcaag gaggagttac ac 22

<210> 114
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 114
ctgcctcaag gaggagttac ac 22

<210> 115
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 115
atttatcccc acgtgagaga gg 22

<210> 116
<211> 21
<212> DNA
<213> Caenorhabditis elegans

<400> 116
cactggatga cagatttgat g 21

<210> 117
<211> 21
<212> DNA
<213> Caenorhabditis elegans

<400> 117
tgatgagaca cgggtgaaac g 21

<210> 118
<211> 21
<212> DNA
<213> Caenorhabditis elegans

<400> 118
gaacggataa aaaggcggag c 21

<210> 119
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 119
ttgatgtgac ctccagatga ac 22

<210> 120
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 120
gcagcacact cttgttttca gc 22

<210> 121
<211> 20
<212> DNA
<213> Caenorhabditis elegans

<400> 121
caaatcactc acttcctgcg 20

<210> 122
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 122
ttcaagtgtc cttgtatccg tg 22

<210> 123
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 123
gcatagaatg gcggaagatc ac 22

<210> 124
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 124
cttccaaatt tgtcctgact gc 22

<210> 125
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 125
aattgcagga gtcgaagttt cc 22

<210> 126
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 126
aacgagcaga caggaaatca tc 22

<210> 127
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 127
tgtgacagca tgtttgaacg tc 22

<210> 128
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 128
agttgtcaag aagtgcgtca ag 22

<210> 129
<211> 20
<212> DNA
<213> Caenorhabditis elegans

<400> 129
gagatggctt gttggacgac 20

<210> 130
<211> 21
<212> DNA
<213> Caenorhabditis elegans

<400> 130
gacaaaatca cgtcacgaag t 21

<210> 131
<211> 21
<212> DNA
<213> Caenorhabditis elegans

<400> 131
ttacttttct gggcagcaag c 21

<210> 132
<211> 30
<212> DNA
<213> Caenorhabditis elegans

<400> 132
cgtgggtatt ccttgttcga agccagctac 30

<210> 133
<211> 23
<212> DNA
<213> Caenorhabditis elegans

<400> 133
tcaagtcaaa tggatgcttg aga 23

<210> 134
<211> 30
<212> DNA
<213> Caenorhabditis elegans

<400> 134
tcacaagctg atcgactcga tgccacgtcg 30

<210> 135
<211> 25
<212> DNA
<213> Caenorhabditis elegans

<400> 135
gattttgtga acactgtggt gaagt 25

<210> 136
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 136
ttattacatc cgtcactgcg tc 22

<210> 137
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 137
gcgtccttat tcagaattcc ag 22

<210> 138
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 138
cttgtgactt caagcccact tc 22

<210> 139
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 139
ggttatgaac cgattaggct cc 22

<210> 140
<211> 21
<212> DNA
<213> Caenorhabditis elegans

<400> 140
gtagccttcc ggggtaaaat c 21

<210> 141
<211> 21
<212> DNA
<213> Caenorhabditis elegans

<400> 141
gatctcgcgc tatgttttga g 21

<210> 142
<211> 21
<212> DNA
<213> Caenorhabditis elegans

<400> 142
gacagctgaa gctgaccaaa c 21

<210> 143
<211> 22
<212> DNA
<213> Caenorhabditis elegans

<400> 143
caggagttaa acgtggtcac tg 22

<210> 144
<211> 31
<212> DNA
<213> Caenorhabditis elegans

<400> 144
cccaagcttg gtttaattac ccaagtttga g 31

<210> 145
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 145
gctctagata attcaatgaa aaggcaaaac gacg 34

<210> 146
<211> 18
<212> DNA
<213> Caenorhabditis elegans

<400> 146
tagaaggcac agtcgagg 18

<210> 147
<211> 19
<212> DNA
<213> Caenorhabditis elegans

<400> 147
taatacgact actataggg 19

<210> 148
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 148
cccaagcttc ttcatttggg cttcatttta ccac 34

<210> 149
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 149
gctctagaga aacaatgttt ttattcaaca tg 32

<210> 150
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 150
ccgctcgagc tcgacgttct tcaatctata tttc 34

<210> 151
<211> 35
<212> DNA
<213> Caenorhabditis elegans

<400> 151
gctctagaca aacaccatta aatctgtatt taaac 35

<210> 152
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 152
ccgctcgagc tcgacgttct tcaatctata tttc 34

<210> 153
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 153
gctctagagt tcacaaattc ataaacaaat acg 33

<210> 154
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 154
cccaagcttg gactttatca caatttccag cac 33

<210> 155
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 155
gctctagagt ttctagattt ttagatttcg tg 32

<210> 156
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 156
ccgctcgaga taatgaagct tcttcttctc attg 34

<210> 157
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 157
gctctagagt ttctagattt ttagatttcg tg 32

<210> 158
<211> 85
<212> PRT
<213> Caenorhabditis elegans

<400> 158
Met Ser Leu His Phe Ser Thr Ile Gln Lys Thr Ile Leu Leu Ile Ser
1 5 10 15

Phe Leu Leu Leu Val Thr Leu Ala Pro Arg Thr Ser Ala Ala Phe Pro
20 25 30

Phe Gln Ile Cys Val Lys Lys Met Glu Lys Met Cys Arg Ile Ile Asn
35 40 45

Pro Glu Gln Cys Ala Gln Val Asn Lys Ile Thr Glu Ile Gly Ala Leu
50 55 60

Thr Asp Cys Cys Thr Gly Leu Cys Ser Trp Glu Glu Ile Arg Ile Ser
65 70 75 80

Cys Cys Ser Val Leu
85

<210> 159
<211> 81
<212> PRT
<213> Caenorhabditis elegans

<400> 159

```
Met Leu Thr His Leu Lys Phe Leu Leu Leu Val Ser Leu Phe Ile Asn
  1               5                  10                  15

Phe Ala Val Ser Ser Glu Asp Ile Lys Cys Asp Ala Lys Phe Ile Ser
             20                  25                  30

Arg Ile Thr Lys Leu Cys Ile His Gly Ile Thr Glu Asp Lys Leu Val
         35                  40                  45

Arg Leu Leu Thr Arg Cys Cys Thr Ser His Cys Ser Lys Ala His Leu
     50                  55                  60

Lys Met Phe Cys Thr Leu Lys Pro His Glu Glu Glu Pro His His Glu
 65                  70                  75                  80

Ile
```

<210> 160
<211> 83
<212> PRT
<213> Caenorhabditis elegans

<400> 160

```
Met Lys Leu Leu Pro Leu Ile Val Val Phe Ala Leu Leu Ala Val Ile
  1               5                  10                  15

Ser Glu Ser Tyr Ser Gly Asn Asp Phe Gln Pro Arg Asp Asn Lys His
             20                  25                  30

His Ser Tyr Arg Ser Cys Gly Glu Ser Leu Ser Arg Arg Val Ala Phe
         35                  40                  45

Leu Cys Asn Gly Gly Ala Ile Gln Thr Glu Ile Leu Arg Ala Leu Asp
     50                  55                  60

Cys Cys Ser Thr Gly Cys Thr Asp Lys Gln Ile Phe Ser Trp Cys Asp
 65                  70                  75                  80

Phe Gln Ile
```

<210> 161
<211> 73
<212> PRT
<213> Caenorhabditis elegans

<400> 161
Met Lys Leu Leu His Ile Phe Ile Ile Phe Leu Leu Phe Gln Ser Cys
1 5 10 15

Ser Asn Lys Met Cys Gln Tyr Ser Lys Lys Lys Tyr Lys Ile Cys Gly
20 25 30

Val Arg Ala Leu Lys His Met Lys Val Tyr Cys Thr Arg Gly Met Thr
35 40 45

Arg Asp Tyr Gly Lys Leu Leu Val Thr Cys Cys Ser Lys Gly Cys Asn
50 55 60

Ala Ile Asp Ile Gln Arg Ile Cys Leu
65 70

<210> 162
<211> 258
<212> DNA
<213> Caenorhabditis elegans

<400> 162
atgtcactgc atttctccac tattcaaaaa acaattcttc taatctcatt cttgctcctc 60
gtaacattgg ctcccagaac aagtgcagct tttccattcc aaatttgtgt caaaaaaatg 120
gaaaaaatgt gcagaatcat caatccagag cagtgtgcac aagtaaataa aatcactgag 180
attggagcat tgacagactg ttgcaccgga ctgtgctcct gggaagaaat ccggatctcc 240
tgctgctccg ttttataa 258

<210> 163
<211> 246
<212> DNA
<213> Caenorhabditis elegans

<400> 163
atgctcacac atctgaaatt cttgcttcta gtgagccttt ttatcaactt cgccgtaagc 60
tctgaagaca tcaaatgcga tgcaaagttc atttcgagaa tcacgaaact ctgtattcac 120
ggaattactg aagataaact tgttcgtctt ctcacaagat gctgcacatc tcactgctcc 180
aaagctcatc tgaaaatgtt ctgcaccctg aaacctcacg aagaagaacc acatcacgaa 240
atctaa 246

<210> 164
<211> 249
<212> DNA
<213> Caenorhabditis elegans

<400> 164

atgaagcttc ttctcattgt ggtttttgct cttttggcag tcatatcaga atcatattct 60
ggaaatgact tccaacctcg tgacaataaa catcattcct atcgttcatg tggggaatcg 120
ttgagccgac gagttgcatt tctgtgtaat ggtggagcta ttcaaacaga aatactaaga 180
gctctggatt gttgttccac tggttgtacg gacaaacaga tcttttcttg gtgtgatttt 240
caaatttga 249

<210> 165
<211> 222
<212> DNA
<213> Caenorhabditis elegans

<400> 165
atgaagcttt tacatatttt tattattttt ctgttattcc aatcgtgctc taataaaatg 60
tgtcaatatt caaagaaaaa gtacaagatt tgtggagtta gagctattaa gcatatgaaa 120
gtctattgta cacgtggaat gacaagagat tatggaaaat tactcgtgac ttgttgttcg 180
aaaggatgta atgcaataga tatccaacgt atttgtttat ga 222

<210> 166
<211> 31
<212> DNA
<213> Caenorhabditis elegans

<400> 166
cccaagcttg gtttaattac ccaagtttga g 31

<210> 167
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 167
gctctagaca attttgatat taaattttgt cg 32

<210> 168
<211> 31
<212> DNA
<213> Caenorhabditis elegans

<400> 168
cccaagcttg gtttaattac ccaagtttga g 31

<210> 169
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 169
gctctagatt aaattttgtc gattttcaag ttg 33

<210> 170
<211> 15
<212> DNA
<213> Caenorhabditis elegans

<400> 170
gttttcccag tcacg 15

<210> 171
<211> 17
<212> DNA
<213> Caenorhabditis elegans

<400> 171
caggaaacag ctatgac 17

<210> 172
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 172
cccaagcttg agcattttgt tgctctgcaa aatg 34

<210> 173
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 173
gctctagatt aaattttgtc gattttcaag ttg 33

<210> 174
<211> 18
<212> DNA
<213> Caenorhabditis elegans

<400> 174
tagaaggcac agtcgagg 18

<210> 175
<211> 19
<212> DNA
<213> Caenorhabditis elegans

<400> 175
taatacgact actataggg 19

<210> 176
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 176
cgggatcccc gcacaaactt atatgacaac tc 32

<210> 177
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 177
cggaattcgg tgtctcataa tggtagtgga tac 33

<210> 178
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 178
cgggatcccc gcacaaactt atatgacaac tc 32

<210> 179
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 179
cggaattcgc aaaagagagg tatagggata aag 33

<210> 180
<211> 18
<212> DNA
<213> Caenorhabditis elegans

<400> 180
tagaaggcac agtcgagg 18

<210> 181
<211> 19
<212> DNA
<213> Caenorhabditis elegans

<400> 181
taatacgact actataggg 19

<210> 182
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 182
cccaagctta aaggcttaga tgcagaaaga cc 32

<210> 183
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 183
gctctagagg gattaaaatc actctgtgat taag 34

<210> 184
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 184
cccaagctta aaggtggaca ttgtagaagg ttg 33

<210> 185
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 185
gctctagagg gattaaaatc actctgtgat taag 34

<210> 186
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 186
cccaagcttc cttcacttct cagcgaagga aatg 34

<210> 187
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 187
gctctagagt gctcatgctc cgttattttg tgc 33

<210> 188
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 188
cggaattcct agaattttca ccccaaatgt tcag 34

<210> 189
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 189
ccgctcgaga aatgtaagtg attggcaagt tgg 33

<210> 190
<211> 33
<212> DNA
<213> Caenorhabditis elegans

<400> 190
cccaagctta gagacttaga cgcaaagagg acc 33

<210> 191
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 191
gctctagagc aggaaaatta gctaaaacat aatg 34

<210> 192
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 192
cggaattcgg cgaaacactt ccgccaactc ac 32

<210> 193
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 193
ccgctcgaga cctaccgtca acttggagga taac 34

<210> 194
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 194
cccaagcttc cttgcacctg ccttcaacca tcac 34

<210> 195
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 195
gctctagata ttctgacccc aaaatgacaa tc 32

<210> 196
<211> 34
<212> DNA
<213> Caenorhabditis elegans

<400> 196
cccaagcttt tctgcagact tgcaaggtta gttc 34

<210> 197
<211> 32
<212> DNA
<213> Caenorhabditis elegans

<400> 197
gctctagaat tcacaaaata atcaagacaa tc 32

<210> 198
<211> 103
<212> PRT
<213> Caenorhabditis elegans

<400> 198

Met	Arg	Ser	Pro	Thr	Leu	Phe	Leu	Leu	Leu	Leu	Leu	Val	Pro	Leu	Ala
1				5				10					15		
Leu	Cys	His	Val	Phe	Ser	Glu	Pro	Ala	Asp	Leu	Glu	Leu	Lys	Ser	Tyr
			20				25					30			
Gln	Ala	Leu	Glu	Lys	Ser	Leu	Lys	Glu	Met	Gly	Leu	Ile	Arg	Ala	Asn
		35				40					45				

```
Gln Gly Pro Gln Lys Ala Cys Gly Arg Ser Met Met Met Lys Val Gln
     50                  55                  60

Lys Leu Cys Ala Gly Gly Cys Thr Ile Gln Asn Asp Asp Leu Thr Ile
 65                  70                  75                  80

Lys Ser Cys Ser Thr Gly Tyr Thr Asp Ala Gly Phe Ile Ser Ala Cys
                 85                  90                  95

Cys Pro Ser Gly Phe Val Phe
            100
```

<210> 199
<211> 72
<212> PRT
<213> Caenorhabditis elegans

<400> 199

```
Met Leu Phe Lys Ile Ile Ile Leu Phe Phe Leu Leu Leu Gln Leu Ser
  1               5                  10                  15

Glu Ala Lys Pro Glu Ala Gln Arg Arg Cys Gly Arg Tyr Leu Ile Arg
             20                  25                  30

Phe Leu Gly Glu Leu Cys Asn Gly Pro Cys Ser Gly Val Ser Ser Val
         35                  40                  45

Asp Ile Ala Thr Ile Ala Cys Ala Thr Ala Val Pro Ile Glu Asp Leu
     50                  55                  60

Lys Asn Met Cys Cys Pro Asn Leu
 65                  70
```

<210> 200
<211> 110
<212> PRT
<213> Caenorhabditis elegans

<400> 200

```
Met Arg Ala Leu Val Ala Ile Leu Cys Leu Met Ala Leu Cys His Ala
  1               5                  10                  15

Ala Met Leu Asp Glu Leu Glu Met Gln Lys Glu Val Gln Glu Phe His
             20                  25                  30

His Met Asn Gly Met Leu Gln Glu Phe Met Asn Lys Gly Leu Ile Gly
```

35 40 45

Asn His His His Gly Thr Lys Ala Gly Leu Thr Cys Gly Met Asn Ile
50 55 60

Ile Glu Arg Val Asp Lys Leu Cys Asn Gly Gln Cys Thr Arg Asn Tyr
65 70 75 80

Asp Ala Leu Val Ile Lys Ser Cys His Arg Gly Val Ser Asp Met Glu
85 90 95

Phe Met Val Ala Cys Cys Pro Thr Met Lys Leu Phe Ile His
100 105 110

<210> 201
<211> 67
<212> PRT
<213> Caenorhabditis elegans

<400> 201
Met Met Arg Ser Phe Phe Val Leu Leu Ala Leu Leu Ala Ile Val Thr
1 5 10 15

Ser Thr Ala Ser Pro Thr Cys Gly Arg Ala Leu Leu His Arg Ile Gln
20 25 30

Ser Val Cys Gly Leu Cys Thr Ile Asp Ala His His Glu Leu Ile Ala
35 40 45

Ile Ala Cys Ser Arg Gly Leu Gly Asp Lys Glu Ile Ile Glu Met Cys
50 55 60

Cys Pro Ile
65

<210> 202
<211> 76
<212> PRT
<213> Caenorhabditis elegans

<400> 202
Met Phe Cys Lys Phe Val Phe Leu Ile Phe Leu Leu Ile Ser Leu Ser
1 5 10 15

Val Ala Thr Ala Asp Phe Gly Ala Gln Arg Arg Cys Gly Arg His Leu
20 25 30

Val Asn Phe Leu Glu Gly Leu Cys Gly Gly Pro Cys Ser Glu Ala Pro
35 40 45

Thr Val Glu Leu Ala Ser Trp Ala Cys Ser Ser Ala Val Ser Ile Gln
50 55 60

Asp Leu Glu Lys Leu Cys Cys Pro Ser Asn Leu Ala
65 70 75

<210> 203
<211> 120
<212> PRT
<213> Caenorhabditis elegans

<400> 203
Met Ser Ser His Ala Leu Val Leu Phe Leu Leu Leu Phe Leu Leu Pro
1 5 10 15

Val Ala Leu Gly His Phe Leu Ser Lys Pro Ala Pro Asp Pro Arg Ile
20 25 30

Thr Phe Asn Arg Lys Leu Ala Glu Thr Leu Lys Glu Leu Gln Asp Met
35 40 45

Gly Leu Ile Gln Ala Pro Arg Glu Pro Val Val Ala Ala Gln Gly Ala
50 55 60

Lys Lys Thr Cys Gly Arg Ser Leu Leu Ile Lys Ile Gln Gln Leu Cys
65 70 75 80

His Gly Ile Cys Thr Val His Ala Asp Asp Leu His Glu Thr Ala Cys
85 90 95

Met Lys Gly Leu Thr Asp Ser Gln Leu Ile Asn Ser Cys Cys Pro Pro
100 105 110

Ile Pro Gln Thr Pro Phe Val Phe
115 120

<210> 204
<211> 218
<212> PRT
<213> Caenorhabditis elegans

<400> 204

```
Met Lys Met Pro Leu Ile Leu Leu Leu Leu Val Ala Ala Ala Ser Ala
 1               5                  10                  15

Phe Val His His Phe Asp His Ser Met Phe Ala Arg Pro Glu Lys Thr
            20                  25                  30

Cys Gly Gly Leu Leu Ile Arg Arg Val Asp Arg Ile Cys Pro Asn Leu
        35                  40                  45

Asn Tyr Thr Tyr Lys Ile Glu Trp Glu Leu Met Asp Asn Cys Cys Glu
    50                  55                  60

Val Val Cys Glu Asp Gln Trp Ile Lys Glu Thr Phe Cys Arg Ala Pro
65                  70                  75                  80

Arg Phe Asn Phe Phe Gly Pro Ser Phe Lys Ala Leu Glu Arg Ser Cys
                85                  90                  95

Gly Pro Lys Leu Phe Thr Arg Val Lys Thr Val Cys Gly Glu Asp Ile
            100                 105                 110

Asn Val Asp Asn Lys Val Lys Ile Ser Asp His Cys Cys Thr Pro Glu
        115                 120                 125

Gly Gly Cys Thr Asp Asp Trp Ile Lys Glu Asn Val Cys Lys Gln Thr
    130                 135                 140

Arg Phe Asn Phe Phe Arg Gln Phe Leu Asp Ser Pro Gln Arg Ser Cys
145                 150                 155                 160

Gly Pro Gln Leu Phe Lys Arg Val Asn Thr Leu Cys Asn Glu Asn Ile
                165                 170                 175

Asn Val Glu Asn Asn Val Ser Val Ser Lys Ser Cys Cys Glu Ser Ala
            180                 185                 190

Ala Gly Cys Thr Asp Asp Trp Ile Lys Lys Asn Val Cys Thr Gln His
        195                 200                 205

Lys Pro Phe Val Phe Arg Pro Gly Phe Tyr
    210                 215
```

<210> 205
<211> 107
<212> PRT
<213> Caenorhabditis elegans

<400> 205
Met Ile Phe Tyr Leu Thr Thr Tyr Leu Val Thr Met Ser Pro Leu Phe
1 5 10 15

Leu Ile Leu Leu Leu Leu Val Ser Thr Thr Tyr Pro Tyr Ile Ile Asp
20 25 30

Ser Ser Glu Ser Tyr Glu Val Leu Met Leu Phe Gly Tyr Lys Arg Thr
35 40 45

Cys Gly Arg Arg Leu Met Asn Arg Ile Asn Arg Val Cys Val Lys Asp
50 55 60

Ile Asp Pro Ala Asp Ile Asp Pro Lys Ile Lys Leu Ser Glu His Cys
65 70 75 80

Cys Ile Lys Gly Cys Thr Asp Gly Trp Ile Lys Lys His Ile Cys Ser
85 90 95

Glu Glu Val Leu Asn Phe Gly Phe Phe Glu Asn
100 105

<210> 206
<211> 77
<212> PRT
<213> Caenorhabditis elegans

<400> 206
Met Gln Ser Leu Pro Ile Leu Ala Cys Leu Leu Thr Leu Ser Val Phe
1 5 10 15

Ala Pro Glu Ile His Gly Arg Glu Leu Lys Arg Cys Ser Val Lys Leu
20 25 30

Phe Asp Ile Leu Ser Val Ile Cys Gly Thr Glu Ser Asp Ala Glu Ile
35 40 45

Leu Gln Lys Val Ala Val Lys Cys Cys Gln Glu Gln Cys Gly Phe Glu
50 55 60

Glu Met Cys Gln His Ala Asn Leu Lys Ile Asp Lys Ile
65 70 75

<210> 207
<211> 312
<212> DNA

<213> Caenorhabditis elegans

<400> 207
atgagatctc ccaccttgtt tcttcttctg ctcctagtgc ccctggcact atgccatgtc 60
ttctcggagc ccgcggattt ggagctcaaa agctaccaag cgcttgaaaa aagcctcaag 120
gagatgggac tcattcgagc caaccaggga cctcaaaaag cgtgcggacg atcaatgatg 180
atgaaggtgc agaagctttg cgcgggcgga tgcacaattc agaacgacga tcttaccatc 240
aaatcctgca gtactgggta caccgatgcc ggcttcatct cggcctgctg cccatctggc 300
ttcgttttct aa 312

<210> 208
<211> 216
<212> DNA
<213> Caenorhabditis elegans

<400> 208
atgttgttca aaatcatcat tttatttttc ctgctccagc tttctgaagc caaaccggaa 60
gcccagaggc gctgcggccg gtatttaatt cgttttttgg gggaactgtg taatggtccc 120
tgctcaggag tttcaagcgt tgacattgcc acaattgcct gtgcaaccgc cgtcccaatc 180
gaagatctga agaatatgtg ttgcccaaat ttgtga 216

<210> 209
<211> 333
<212> DNA
<213> Caenorhabditis elegans

<400> 209
atgagagctc tcgtcgctat tctctgcctt atggcactat gccatgcagc aatgctcgat 60
gagctggaga tgcagaagga ggttcaggag ttccatcaca tgaacggcat gctccaagag 120
ttcatgaata aggggctcat cgggaatcat caccatggta ccaaggccgg cctcacctgc 180
gggatgaaca tcatcgagag agtcgacaag ctgtgcaatg ggcagtgcac tcggaactat 240
gatgcactcg tcatcaagtc ctgccaccgc ggagtctcgg acatggagtt catggtggca 300
tgctgcccaa ccatgaagct attcattcac taa 333

<210> 210
<211> 204
<212> DNA
<213> Caenorhabditis elegans

<400> 210
atgatgcgct cattctttgt gctcttggct ctgctcgcaa tagtcaccag caccgctagt 60
cccacttgtg gcagggctct tctacaccgg atccagtcgg tttgcggtct ctgtaccatc 120
gacgctcacc acgaactgat tgccattgcc tgctcaaggg gactgggcga taaggaaatc 180
attgaaatgt gctgtccaat ctaa 204

<210> 211
<211> 231
<212> DNA

<213> Caenorhabditis elegans

<400> 211
atgttctgta aatttgtatt cctgatcttt ctactcatct ctctgtcagt ggccaccgct 60
gactttggcg cccagcgccg ttgtgggcgc cacttggtga acttcctcga gggactctgc 120
ggtggcccgt gctctgaagc tccgactgtt gaactagctt cgtgggcatg ttcatcagca 180
gtctcaattc aggatctcga aaaattgtgc tgtccttcaa atcttgcttg a 231

<210> 212
<211> 363
<212> DNA
<213> Caenorhabditis elegans

<400> 212
atgagttctc acgccctggt tcttttcctt ctccttttcc tcctaccagt ggcactgggc 60
cacttcctct ccaagcctgc accggatcca aggatcacat tcaaccgtaa gcttgcggag 120
acactcaagg agcttcagga catgggactc atccaggccc cccgtgagcc ggtagtggcg 180
gctcagggag ccaagaagac ttgcggaagg agtttgttga taaagatcca acaactctgc 240
catggaatct gcacagttca cgctgatgac ctccacgaaa cggcatgcat gaaaggtctc 300
accgactctc agctgatcaa ctcctgctgc ccaccaatcc cccagacacc attcgtcttc 360
tga 363

<210> 213
<211> 657
<212> DNA
<213> Caenorhabditis elegans

<400> 213
atgaagatgc ccttgatctt gctgcttctc gtcgccgccg catcggcgtt cgtccaccac 60
tttgaccatt caatgtttgc cagaccggag aaaacgtgtg gaggactact cattcgtcgt 120
gtcgatagaa tttgcccgaa tctaaattat acatataaaa ttgactggga acttatggac 180
aactgttgcg aagtggtttg cgaggaccag tggattaagg aaacctttg cagagcgccc 240
aggttcaact ttttcggacc ttcattcaaa gcccttgaaa gatcgtgtgg accaaaactg 300
ttcacaaggg ttaaaactgt gtgcggtgaa gacatcaatg ttgataataa agtcaagatt 360
tcggatcact gctgcacacc agagggagga tgcacagacg actggatcaa ggagaacgtc 420
tgcaaacaga ccagattcaa ctttttccga caatttctcg attcccctca aagatcatgt 480
ggaccccagt tgttcaaaag agtgaatact ttgtgtaatg aaaatatcaa tgttgaaaat 540
aatgtaagcg tgtcgaaaag ctgttgcgaa tcagcggcag gatgcacgga tgattggatt 600
aagaagaatg tctgcacaca gcataagcct tttgttttcc gtccagcctt ttactga 657

<210> 214
<211> 324
<212> DNA
<213> Caenorhabditis elegans

<400> 214
atgattttct atctgacaac ctacctagta actatgtcac ctctcttcct gatcctgttg 60
cttctagtct ctaccactta cccttacatc attgactctt cggagagtta tgaagttcta 120

atgctattcg ggtataagag aacatgtgga cgacgcttga tgaacaggat taatagagta 180
tgcgtgaagg atatagatcc agcagatatc gatccgaaga tcaaattatc ggagcactgt 240
tgtatcaagg gatgcacaga tggatggatc aagaagcata tttgcagtga ggaagttctg 300
aattttggat tttttgaaaa ttga 324

<210> 215
<211> 234
<212> DNA
<213> Caenorhabditis elegans

<400> 215
atgcaaagcc taccaattct tgcctgcctc ctcacactgt cagtttttgc gccggaaatt 60
catggccggg agctcaaacg ttgttctgtg aaacttttttg atattctaag cgtaatttgt 120
ggaactgaaa gtgatgcaga aattctacaa aaagtcgcag tgaaatgctg ccaggagcag 180
tgtgggtttg aggaaatgtg ccagcatgcc aacttgaaaa tcgacaaaat ttaa 234

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/04, C07K 14/00, 16/00, C12N 15/00 | A3

(11) International Publication Number: **WO 99/54436**

(43) International Publication Date: 28 October 1999 (28.10.99)

(21) International Application Number: PCT/US99/08522

(22) International Filing Date: 16 April 1999 (16.04.99)

(30) Priority Data:

09/062,580	17 April 1998 (17.04.98)	US
09/074,984	8 May 1998 (08.05.98)	US
09/084,303	26 May 1998 (26.05.98)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	09/084,303 (CIP)
Filed on	26 May 1998 (26.05.98)
US	09/062,580 (CIP)
Filed on	17 April 1998 (17.04.98)
US	09/074,984 (CIP)
Filed on	8 May 1998 (08.05.98)

(71) Applicant *(for all designated States except US)*: EXELIXIS PHARMACEUTICALS, INC. [US/US]; 260 Littlefield Avenue, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: HOMBURGER, Sheila, A. [US/US]; 3974 25th Street, San Francisco, CA 94114 (US). PLATT, Darren, M. [US/US]; 929 Pine Street, San Francisco, CA 94108 (US). FERGUSON, Kimberly, C. [US/US]; 1038 Chula Vista Avenue, Burlingame, CA 94010 (US). DOBERSTEIN, Stephen, K. [US/US]; 2 Townsend Street, San Francisco, CA 94107 (US). BUCHMAN, Andrew, R. [US/US]; 3119 Eton Avenue, Berkeley, CA 94705 (US). REDDY, Bindu, P. [IN/US]; 855 Bounty Drive, Foster City, CA 94404 (US).

(74) Agents: ANTLER, Adriane, M. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report:
29 December 1999 (29.12.99)

(54) Title: NUCLEIC ACIDS AND PROTEINS OF *C. ELEGANS* INSULIN-LIKE GENES AND USES THEREOF

(57) Abstract

The present invention relates to *C. elegans* insulin–like genes and methods for identifying insulin–like genes. The methods provide nucleotide sequences of *C. elegans* insuline–like genes, amino acid sequences of their encoded proteins, and derivatives (e.g., fragments) and analogs thereof. The invention further relates to fragments (and derivatives and analogs thereof) of insuline–like proteins which comprise one or more domains of an insulin–like protein. Antibodies to an insuline–like protein, and derivatives and analogs thereof, are provided. Methods of production of an insuline–like protein (e.g., by recombinant means), and derivatives and analogs thereof, are provided. Further, methods to identify the biological function of a *C. elegans* insulin–like gene are provided, including various methods for the functional modification (e.g., overexpression, underexpression, mutation, knock–out) of one or more genes silmutaneously. Still further, methods to identify a *C. elegans* gene which modifies the function of, and/or functions in a downstream pathway from, an insulin–like gene are provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/08522

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/04; C07K 14/00, 16/00; C12N 15/00
US CL :530/350, 387.1; 535/23.5; 800/13

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350, 387.1; 535/23.5; 800/13

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LIN et al. *daf-16*: An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. 14 November 1997, Vol. 278, pages 1319-1322, entire document.	1-40
A	OGG et al. The fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. Nature. 30 October 1997, Vol. 389, pages 994-999, entire document.	1-40

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
12 SEPTEMBER 1999

Date of mailing of the international search report
25 OCT 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
Anne-Marie Baker, Ph.D.
Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/08522

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS
DIALOG (file: medicine)
search terms: C. elegans, insulin, transgen?, antibody, chimeric, nucleotide